(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/79264 A1

(51) International Patent Classification[7]: G01N 33/00, A01K 67/00, 67/033, C12N 15/00, 15/09, 15/63, 15/70, 15/74, 5/00, 5/02, C07H 21/02, 21/04

(21) International Application Number: PCT/US00/17493

(22) International Filing Date: 23 June 2000 (23.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/140,920 24 June 1999 (24.06.1999) US

(71) Applicant *(for all designated States except US)*: **BAYLOR COLLEGE OF MEDICINE** [US/US]; One Baylor Plaza, Suite 186A, Houston, TX 77030 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BEAUDET, Arthur, L.** [US/US]; 48 Stillforest, Houston, TX 77024 (US). **TSAI, Ting-Fen** [—/—]; 20 Tung-Hwa Street (3F), Lane 174, Section 2, Pei-Tou 112, Taipei (TW). **CHEN, Ken-Shiung** [—/US]; 3607 Austin Meadow Court, Sugar Land, TX 77479 (US). **WEBER, John, S.** [US/US]; 3327 East Overdale, Pearland, TX 77584 (US). **JUSTICE, Monica, J.** [US/US]; 5011 Yarwall, Houston, TX 77096 (US). **ZOGHBI, Huda, Y.** [US/US]; 6618 Sewanee, Houston, TX 77005 (US).

(74) Agents: **PAUL, Thomas, D.** et al.; Fulbright & Jaworski, L.L.P., 1301 McKinney, Suite 5100, Houston, TX 77010 (US).

(81) Designated States *(national)*: AU, CA, JP, MX, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/79264 A1

(54) Title: SYSTEM AND METHOD TO DETECT COMPOUNDS AND MUTATIONS THAT UPREGULATE AND DOWN-REGULATE EXPRESSION OF NUCLEOTIDE SEQUENCES

(57) Abstract: This invention relates to the use of a system in which expression of a reporter gene is indicative of the activity of a promoter by which it is regulated. More particularly it relates to administering a compound and assaying subsequent changes in said promoter activity through expression of said reporter gene which resides within the nucleotide sequence being regulated by said promoter. Mutagenesis of cells which contain said reporter gene system allows isolation of upstream regulatory mutations or modifier mutations.

System and Method to Detect Compounds and Mutations that Upregulate and Downregulate Expression of Nucleotide Sequences

This application claims priority to a U.S. provisional application 60/140,920 filed June 24, 1999.

The work herein was supported by grants from the United States Government. The United States Government may have certain rights in the invention

FIELD OF THE INVENTION

This invention relates to the use of a reporter gene system to detect alterations in the activity of a promoter. More particularly it relates to the use of said reporter gene system to screen for mutations and compounds which upregulate or downregulate a particular nucleotide sequence expression.

BACKGROUND OF THE INVENTION

The molecular genetics of genomic imprinting can be studied utilizing the homologous regions of mouse chromosome 7C and human chromosome 15q11-q13 that include the genes responsible for Prader-Willi syndrome (PWS) and Angelman syndrome (AS) in the human. PWS and AS are distinct neurogenetic disorders caused by deficiency of paternal or maternal gene expression respectively from human chromosome 15q11-q13 (Jiang et al., 1998; Nicholls et al., 1998). AS is caused by maternal deficiency for E6-AP ubiquitin-protein ligase (Malzac et al., 1998; Fang et al., 1999), but the genes causing PWS are not known. The human *SNRPN* and mouse *Snrpn* loci are within the PWS critical region and are expressed exclusively from the paternal chromosome and encode a small protein of unknown significance SNURF and small nuclear ribonucleoprotein-associated polypeptide N (SmN) (Ozçelik et al., 1992; Leff et al., 1992; Reed and Leff, 1994). The *SNRPN/Snrpn* loci in human and mouse are associated with a CpG island that encompasses the promoter and is heavily methylated exclusively on the maternal, silenced chromosome. *Snrpn* is known to be ubiquitously expressed in the mouse with abundant expression in brain and

heart and lower levels of expression in a wide range of tissues in both adults and developing embryos (Barr et al., 1995). All available evidence is consistent with very tight silencing of *SNRPN/Snrpn* on the maternal chromosome in human and mouse in all tissues and at all developmental stages studied to date (Barr et al., 1995). Furthermore, mice lacking *Snrpn* expression on the basis of maternal uniparental disomy show normal regulation of neuronal alternative splicing events (Huntriss et al., 1994). There is evidence based on disruption of *SNRPN* in two patients with PWS (Sun et al., 1996; Kuslich et al., 1999) that paternal deficiency of *SNRPN* may cause PWS, but there are also rare patients suggesting exclusion of *SNRPN* as a causative gene in PWS (Schulze et al., 1996; Conroy et al., 1997). However mice with maternal uniparental disomy (UPD) causing paternal deficiency for chromosome 7C (Cattanach et al., 1992) or with an imprinting mutation causing paternal deficiency of *Snrpn* and the surrounding region (Yang et al., 1998) demonstrate postnatal lethality that has been proposed to be the equivalent of the neonatal PWS phenotype, but mice with a heterozygosity or homozygosity for a null mutation for *Snrpn* do not show lethality or any other obvious phenotype (Yang et al., 1998).

There is evidence for both *cis* elements and *trans*-acting factors in the regulation of imprinted gene expression (Surani, 1998; Reik and Walter, 1998). In particular, there is evidence for a bipartite imprinting center (IC) at the site of the *SNRPN* promoter and further upstream in the human, with a more proximal region required for switching from maternal to paternal epigenotype and a more upstream region required for switching from paternal to maternal epigenotype (Dittrich et al., 1996). More generally, there are extensive data regarding the role of promoters, enhancers, and a genomic mark in the imprinted expression of *H19* and *Igf2* in the mouse (Bartolomei and Tilghman, 1997). Numerous *trans*-acting factors must be required for the switching, establishment, and maintenance of genomic imprinting. For example, absence of the DNA methyltransferase in mice causes lethality and abnormalities of imprinting (Li et al., 1992; Li et al., 1993; Warnecke et al., 1998). The interaction of *cis* elements with *trans*-acting factors may establish a hierarchical control mechanism for the initiation and maintenance of genomic imprinting, and mutations might affect various aspects of this mechanism. In some instances, *cis* elements taken from imprinted regions in the mouse and human have been shown to have silencing effects in *Drosophila* (Lyko et al., 1997; Lyko et al., 1998).

By expressing coat color markers under the control of various promoters, this approach could provide a general strategy for screening for regulatory mutations using ethylnitrosourea (ENU) or insertional mutagenesis (Justice et al., 1997; Schimenti and Bucan, 1998; Woychik and Alagramam, 1998; Battey et al., 1999). Genetic screens based on easily visible markers in *Drosophila* have been the cornerstone for dissecting genetic and epigenetic (e.g., transfection and position-effect variegation) regulatory mechanisms. The strategy of genetic fusions as experimental tools is well established in bacteria (Slauch and Silhavy, 1991).

Although about 100 genes in the mouse have been identified as affecting coat color (Barsh, 1996; Jackson, 1994), the agouti locus is particularly attractive for use in these studies because the genetics are well characterized, and phenotypic effects are known over a wide range of expression. The agouti gene on mouse chromosome 2 encodes a 131 amino acid secreted protein(Bultman et al., 1992; Miller et al., 1993) that is an antagonist of the melanocyte-stimulating-hormone (MSH) receptor (Lu et al., 1994). The agouti protein acts in a paracrine manner to inhibit stimulation of melanogenesis by αMSH and induces melanocytes to switch from producing black pigment (eumelanin) to yellow pigment (phaeomelanin) (Silvers, 1979). At least 19 dominant and recessive alleles have been identified at the agouti locus (Green, 1989). The most common alleles in laboratory mice are agouti (*A*), which is dominant and yields a brown coat color, and the recessive nonagouti (*a*) allele, which yields a black color in homozygotes. The agouti locus is also attractive for study because extremely low levels of expression yield a phenotypically significant change in coat color compared to homozygotes for the extreme nonagouti allele (a^e), a null mutation (Hustad et al., 1995). It is predicted that even low levels of expression of the protein encoded by the *A* allele might yield visible changes in coat color.

In mice, agouti is expressed only in the skin during the time of phaeomelanin synthesis and not in brain, liver, kidney, or numerous other tissues (Bultman et al., 1992; Miller et al., 1993). However, several dominant mutations of the agouti locus or ectopic production controlled by a transgene with a heterologous promoter are associated with increased yellow pigment in the coat, obesity, diabetes, and susceptibility to tumors (Duhl et al., 1994a; Duhl et al., 1994b; Wolff et al., 1987). Furthermore, the ubiquitous overexpression of agouti under the control of the β-actin or Pgk-1 promoter in transgenic

mice also gives rise to yellow, obese mice (Klebig et al., 1995; Perry et al., 1995). Michaud et al. described a new dominant agouti allele designated A^{iapy} (Michaud et al., 1994). This mutation arose through the spontaneous insertion of an intracisternal A-particle (IAP) proviral element 51-bp upstream of the first coding exon of the agouti locus. These authors demonstrated that ectopic expression of the agouti gene arose through the utilization of a cryptic promoter within the 5′ long terminal repeat (LTR) of the provirus. In addition, the methylation status of the 5′ LTR was inversely correlated with the level of A^{iapy} expression; most importantly, methylation and expression varied with maternal versus paternal inheritance (Michaud et al., 1994). This nonphysiological "imprinting" of the A^{iapy} allele provided a stimulus to explore the strategy of expressing the agouti cDNA under the control of an imprinted promoter. The coat color of mice with the A^{iapy} allele ranges from solid yellow to pseudoagouti, the latter being similar to the *A/A* genotype. Mice with high levels of ectopically expressed agouti mRNA show yellow coat color and marked obesity, whereas pseudoagouti mice express the agouti gene ectopically at low levels, and their body weights are similar to controls (Michaud et al., 1994)

There is a general appreciation in the research community that better methods are needed for mutagenesis and genetic screening in mammalian organisms such as the mouse and that better methods are needed for screening for compounds that affect gene expression. As in the example of the Big Blue mouse system advocated by Stratagene (La Jolla, CA), these are very cumbersome systems by comparison to that which is proposed in this patent application. The intensity of scientific interest in the mouse is well summarized in a recent article entitled "An action plan for mouse genomics" (Battey *et al.*, 1999). The intensity of scientific interest is also indicated by two RFA's from the NIH: HD-99-007 entitled "Mouse mutagenesis and phenotyping: Developmental defects" and MH-99-007 entitled "Mouse mutagenesis and phenotyping: Nervous system and behavior." In general, molecular genetic manipulations possible in the mouse can ordinarily be carried out in other mammalian species and lower organisms once the feasibility is demonstrated in the mouse. This patent application proposes a general method to screen for genetic effects and/or effects of various compounds on gene expression, both transcriptional and posttranscriptional.

This invention demonstrates for the first time the use of knockin/fusions in higher eukaryotic organisms to detect compounds and mutations that upregulate and downregulate

expression of a specific gene. It is noteworthy that although similar methods in bacteria have existed for nearly a decade, the idea and use of a similar general strategy for such detection in mammals or other organisms has gone unrealized, suggesting that this invention is indeed nonobvious.

SUMMARY OF THE INVENTION

An embodiment of the present invention is a system to determine elements that regulate expression of a nucleotide sequence comprising a reporter sequence inserted within said nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for said nucleotide sequence. In specific embodiments, the reporter sequence can be inserted into the part of the sequence coding for the N-terminal portion of a protein, the C-terminal portion of a protein or the 3′ untranslated region of the nucleotide sequence.

In a preferred embodiment of the present invention two different reporter sequences are inserted within the nucleotide sequence and both reporter sequences are controlled by the promoter for the nucleotide sequence.

In specific embodiments of the present invention the nucleotide sequence of the system is selected from the group of organisms consisting of mammals, non-mammalian animals, plants, insects, aquatic organisms and avian species.

An additional embodiment includes a composition of matter wherein a functional system is inserted into the genome of a mammal, a non-mammalian animal, a plant, an insect, an aquatic organism, or an avian species. Additionally a functional system can be inserted into a cell.

An additional embodiment of the present invention is the method of identifying an upregulator or downregulator of expression of a specific nucleotide sequence comprising the steps of inserting a test system into a test organism wherein the test system comprises a reporter sequence inserted into a specific nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for the nucleotide sequence; delivering a compound to the test organism containing the test system; and measuring the effect of said compound on said reporter sequence wherein when said

compound increases the expression of said reporter sequence it is an upregulator and when said compound decreases the expression of said reporter sequence it is a downregulator.

In a specific embodiment the method can be used to identify compounds that affect postranscriptional control at the RNA level when the reporter is inserted by homologous recombination and the resulting construct expresses a fusion transcript.

In an additional specific embodiment, the method can be used to identify compounds that affect posttranscriptional control at the protein level by inserting the reporter sequence into the nucleotide sequence by homologous recombination and such construct expresses a fusion protein

In specific embodiments, the method utilizes a test organism selected from the group consisting of a mammal, a non-mammalian animal, a plant, an insect, an aquatic organism, and an avian species .

A further embodiment of the present invention is the method for identifying an upregulator or downregulator of expression of a specific nucleotide sequence comprising the steps of inserting a test system into a test organism, wherein the test system comprises a reporter sequence inserted within the specific nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for the specific nucleotide sequence, mutagenizing the test organism containing the test system, and measuring the effect of said mutagenesis on expression of said reporter sequence wherein when the mutant resulting from said mutagenizing increases the expression of the reporter sequence, the mutant is an upregulator and when the mutant resulting from said mutagenizing decreases the expression of said reporter sequence, the mutant is a downregulator.

In a specific embodiment, the method further comprises the steps of delivering a compound to the mutagenized test organism and measuring the effect of said compound on said reporter sequence, wherein when said compound increases the expression of said reporter sequence the compound is an upregulator and when said compound decreases the expression of said reporter sequence the compound is a downregulator.

In a preferred embodiment, the method uses two different reporter sequences inserted within the nucleotide sequence and both are controlled by the same promoter.

A further embodiment of the present invention is the method of identifying an upregulator or a downregulator to treat a disease in which there is a deficiency in expression (upregulator) or excess in expression (downregulator) of a nucleotide sequence.

An additional embodiment of the present invention is the method of treating a disease in an organism in which the disease is the result of a deficiency or excess in expression of a nucleotide sequence comprising the step of introducing a pharmacologically effective dose of an upregulator (deficiency) or downregulator (excess).

A further embodiment of the present invention is the method to identify an upregulator or downregulator to enhance food content of an animal, aquatic organism, avian species, or plant.

Other and further objects, features and advantages would be apparent and eventually more readily understood by reading the following specification and by reference to the company drawing forming a part thereof, or any examples of the presently preferred embodiments of the invention are given for the purpose of the disclosure.

DESCRIPTION OF THE DRAWINGS

Figures 1a through 1b show generation of agouti knockin mice by homologous recombination. Figure 1a illustrates the structure of the targeting vector and partial restriction map of the *Snrpn* wild type allele and agouti knockin allele after homologous recombination. Differentially methylated sites of *Snrpn* intron 1 detected by methylation sensitive enzymes *Bss*HII and *Sac*II are shown below the line. The 5′ and 3′ flanking probes are indicated. Restriction enzymes and other designations are as follows: B, *Bss*HII; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I; X, *Xba*I; Sc, *Sac*II; Ag, agouti cDNA; pA, rabbit β-globin polyadenylation signal; Neo, neomycin expression cassette; TK, HSV-tk expression cassette; and Me probe, probe for methylation study presented in Figure 2. ATG^1 represents the potential start site for a small (SNURF) open reading frame in the *Snrpn* gene of uncertain significance. ATG^2 is the agouti start codon and is not in frame with ATG^1, ATG^3 is the start codon for the major (SmN) open reading frame for *Snrpn*. Figure 1b is a Southern blot analysis of ES cell DNA and mouse tail DNA from heterozygous matings using 3′ flanking probe. The *Pst*I 9-kb fragment was detected in the wild type allele (WT), and a *Pst*I

5.2-kb fragment was detected in the agouti knockin allele (Ag KI). Genotypes are indicated as wild type (+) or transgenic (T).

Figures 2a through 2b show a methylation study of the *Snrpn* CpG island in three generations of mice carrying the *Snrpn*/agouti knockin/fusion sequence. Figure 2a depicts breeding of the mice with agouti knockin/fusion sequence expressed from *Snrpn* promoter. Symbols are as follows: squares, males; circles, females; checkered symbols, chimera; filled symbols, mice expressing transgene; symbols with central dots, mice with transgene but not expressed. Figure 2b shows genomic DNA from mice I-1, II-2, and III-2 indicated in panel a was digested with *Hin*dIII (H) alone, *Hin*dIII plus *Sac*II (Sc), or *Hin*dIII plus *Bss*HII (Bs), and analyzed by Southern blot hybridization. Probe for methylation analysis is indicated in Fig.1a (Me probe). Detected fragment sizes are as follows: methylated allele of wild type *Snrpn* (WT), 14 kb; methylated allele of agouti knockin (Ag KI), 12 kb; unmethylated allele of wild type *Snrpn* (WT unme), 11 kb; unmethylated allele of agouti knockin (Ag KI unme), 8.5 kb for *Hin*dIII plus *Sac*II digestion, and 8 kb for *Hin*dIII plus BssHII digestion.

Figures 3a through 3b demonstrate the effect of the transgene on coat color. Photographs of mice with agouti transgene expressed from the endogenous *Snrpn* promoter on a nonagouti background. Figure 3a shows that when the agouti transgene was inherited from a heterozygous male, the offspring carrying the paternally inherited agouti transgene (T^p/+) showed a distinctly lighter tannish coloration on their ventrum compared to littermates without the transgene (+/+). Figure 3b shows that when the agouti transgene was inherited from a heterozygous female, the offspring (T^m/+) were entirely black and were indistinguishable from littermates not carrying the transgene.

Figures 4a through 4b show expression of *Snrpn* and *Snrpn*/agouti fusion transcripts in the skin of the agouti mice with the knockin/fusion sequence during three generations of breeding. Figure 4a indicate partial structure of the *Snrpn* wild type and knockin/fusion sequence. Symbols are the same as in Figure1a. Arrows above the boxes indicate transcriptional orientation of the *Snrpn* and neomycin genes. Arrowheads below the exons indicate primers used for RT-PCR analysis. Primer p3 was used for RT reaction of *Snrpn* RNA, primers p1 and p2 were for PCR of the *Snrpn* cDNA. Primer p5 was used for RT reaction of *Snrpn*/agouti fusion RNA, and primers p1 and p4 were for the subsequent PCR of the *Snrpn*/agouti fusion cDNA. Figure 4b shows RT-PCR analysis was performed for

Snrpn, *Snrpn*/agouti fusion, neomycin, and an internal control gene, *Hprt*, using total RNA isolated from skin of mice I-1, II-2, and III-1 as indicated in the pedigree in Figure 2a. Lanes representing RNA incubated with reverse transcriptase are indicated as (+), while (-) lanes are reactions incubated without reverse transcriptase as controls for genomic DNA contamination.

Figures 5a through 5b illustrate the strategy for detecting dominant mutations affecting genomic imprinting. In Figure 5a mutagenized males homozygous for the transgene and nonagouti *a/a* are bred to nontransgenic nonagouti *a/a* females. In Figure 5b mutagenized nontransgenic nonagouti *a/a* males are bred to females homozygous for the transgene and nonagouti *a/a*.

Figure 6 illustrates a mutant mouse with tannish abdomen compared to nonmutant littermate. This mutant animal was discovered in the screen mating a mutagenized wild type male with a homozygous transgenic female. All offspring in this screen have a black abdomen except for the one mutant animal depicted.

Figure 7 shows a Southern blot demonstrating lack of methylation of the maternally inherited knockin/fusion gene in the mutant mouse. Southern blotting was performed using the same probe and restriction enzyme combination as for Figure 2.

Figure 8 shows preparation of bicistronic agouti and GFP reporter under control of the *Snrpn* promoter. The agouti (Ag) cDNA was placed under the control of the *Snrpn* promoter as for the construct shown in Figure 1.

Figure 9 depicts generation of two fusion gene constructs expressing human α-fetoprotein under control of the promoter for the LDL receptor.

DETAILED DESCRIPTION OF THE INVENTION

It is readily apparent to one skilled in the art that various embodiments and modifications may be made to the invention disclosed in this Application without departing from the scope and spirit of the invention.

The term "aquatic organism" as used herein is defined as freshwater or saltwater organisms. Examples would be fish, shrimp, scallops, mussels and other fresh or salt water food organisms.

The term "compound" means any protein, peptide, nucleic acid, DNA, RNA, oligonucleotide, hormone, sugar, lipid and other inorganic and organic chemical including metals or combinations thereof.

The term "DNA" as used herein is defined as deoxyribonucleic acid.

The term "downregulator" as used herein is defined as a compound or molecule which decreases expression of a particular nucleotide sequence. The decreased expression means less expression than the cell or organism is expressing without the downregulator. This means there is less expression than is present in the native cell, which can be measured before the administration of the downregulator. A downregulator can be any compound or mutant sequence.

The term "elements" as used herein is defined as a compound, mutagen, or any other entity which causes an effect on regulation of expression of said nucleotide sequence.

The term "expression" as used herein is defined as the transcription and/or translation of a particular nucleotide sequence driven by its promoter.

The term "fusion gene" as used herein is used interchangeably with the term "knockin" or "knockin/fusion" and is defined as an insertion of a first nucleic acid into a second nucleic acid, wherein said insertion creates a novel DNA joint. Thus, the first and second nucleic acids are therein made contiguous, and in specific embodiments the translational and/or transcriptional signals which affect one, affect the other. In a preferred embodiment a first nucleic acid is an easily detectable reporter gene or bicistronic reporter cassette. In another preferred embodiment the second nucleic acid is a nucleic acid of interest, the promoter of which, upon creation of the fusion gene, regulates expression of the reporter gene or bicistronic reporter cassette. In another preferred embodiment creation of the fusion gene occurs through homologous recombination.

The term "homologous recombination" as used herein is defined as an exchange between any pair of DNA sequences having a similar sequence of nucleotides, where the two sequences interact (recombine) to form a new recombinant DNA species. The frequency of homologous recombination increases as the length of the shared nucleotide

DNA sequences increase. Homologous recombination can occur between two DNA sequences that are less than identical, but the recombination frequency declines as the divergence between the two sequences increases.

The term "knockin" as used herein is used interchangeably with the term "fusion gene" or the act of creating a "fusion gene". This term is also used interchangeable with "knockin/fusion."

The term "peptide" as used herein is defined as a molecule of at least 10 amino acids in length.

A compound or composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient mammal. Such an agent is said to be administered in a "pharmacologically effective dose" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in technical change in the physiology of a recipient mammal. For example, in the treatment of a disease or condition, a compound which inhibits tumor growth or decreases the size of the tumor would be therapeutically effective; a compound which slows the progression of the disease, prevents the onset, delays the onset or completely treats the disease, would be considered therapeutically effective.

The term "promoter" as used herein is defined as the region of nucleotide sequence which regulates transcription of a specific nucleotide sequence. The term promoter includes enhancers, silencers and other *cis*-acting regulatory elements.

The term "RNA" as used herein is defined as ribonucleic acid.

The term "reporter sequence" as used herein is defined as the nucleotide sequence which when expressed can be detected. The expressed product itself can be detected, such as an RNA or protein, or a metabolite or other characteristic secondarily affected by the reporter product can be detected. The skilled artisan recognizes that any reporter gene that could be detected by transcutaneous monitoring, by visualization with UV light, by visualization with infrared light, or by visualization with other imaging techniques, such as X-ray or MRI, would be of obvious value. Any tissue or body fluid or cell culture or cell free extract can be sampled depending on the marker used. For example, coat color, hair color, skin quality, eye color, secreted proteins, histological markers and other markers used by those skilled in the art.

The term "transcription" as used herein is defined as the synthesis of an RNA directed by a DNA template.

The term "translation" as used herein is defined as the synthesis of a peptide directed by an RNA template.

The term "upregulator" as used herein is defined as a compound or molecule which increases expression of a particular nucleotide sequence. The increased expression means greater expression than the cell or organism is expressing without the upregulator. This means there is more expression than is present in the native cell, which can be measured before the administration of the upregulator. An upregulator can be any compound or mutant sequence.

One specific embodiment of the present invention is a system to determine elements that regulate expression of a nucleotide sequence comprising a reporter sequence inserted within said nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for said nucleotide sequence. A skilled artisan has available the sequences which may be utilized in the present invention as described herein or as is readily available at nationally maintained sequence repositories such as GenBank (http://www.ncbi.nlm.nih.gov/Entrez/index.html). Similarly, a skilled artisan may obtain cell lines and other biological materials through an archive such as the American Type Cell Culture (ATCC) (http://phage.atcc.org/searchengine/all.html) or through other similar repositories. Also, a skilled artisan has available databases such as are available at http://www.ornl.gov/TechResources/Trans/hmepg.html or http://tbase.jax.org/, in which over 800 records of animal lines are maintained, with such detailed information as the methods for the generation of each line, a detailed description of the DNA construct used, expression, phenotype, *etc.*

Another specific embodiment of the present invention is a method to identify an upregulator or downregulator of expression for a specific nucleotide sequence comprising the steps of inserting a test system into a test organism, wherein the test system comprises a reporter sequence inserted into a specific nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for the specific nucleotide sequence, delivering a compound, and measuring the effect of the compound on

the reporter sequence, wherein when said compound increases the expression of said reporter sequence, the compound is an upregulator and when said compound decreases the expression of the reporter sequence, the compound is a downregulator.

Another specific embodiment of the present invention is a method to identify an upregulator or downregulator of expression of a specific nucleotide sequence comprising the steps of inserting a test system into a test organism, wherein the test system comprises a reporter sequence inserted into a specific nucleotide sequence under conditions, wherein the expression of the reporter sequence is controlled by the promoter, mutagenizing the test organism containing the test system and measuring the effect of the mutagenesis on expression of the reporter sequence expression wherein, when the mutant resulting from said mutagenizing increases the expression of the reporter sequence, the mutant is an upregulator and when the mutant resulting from said mutagenizing decreases the expression of said reporter sequence, the mutant is a downregulator. In a further embodiment a test compound is delivered to the mutagenized test organism and its effects on regulation of reporter sequence expression is similarly measured.

In preferred embodiments of the present invention the reporter sequence is two different reporter sequences which are inserted into the specific nucleotide sequence and both are controlled by the promoter of the specific nucleotide sequence. One skilled in the art will recognize specific advantages of using two different reporters in removing the effects of randomness, mutations, etc. Also one of the reporters can be a quick screen, like hair or eye color and the other could be a chemical screen like CAT, luciferase, etc.

Specific embodiments also include composition of matter wherein a system as described herein is functional and is inserted into the genome of an organism selected from the group consisting of non-human mammal, non-mammalian animal, plant, insect, aquatic organism and an avian species. In other specific embodiments the method described herein can use the same above group of organisms as a test organism.

Another specific embodiment includes a cell, cell line, or cell free extract wherein a functional system as described herein is inserted. The cell, cell line, or cell free extract can be used as the test organism in any of the methods described herein. In addition to the organisms described above, human cell lines are also useful. In a specific embodiment

tissue culture cells are utilized for drug or genetic screening through methods of the present invention.

In a preferred embodiment a fusion gene is created and introduced into cultured mouse ES cells, followed by transmission of the mouse ES alterations to the germline by methods well known in the art, and cultured human cells. In a specific embodiment a reporter sequence is inserted, such as by homologous recombination, into a nucleic acid of interest. Such an insertion may or may not disrupt the gene of interest to the extent that a functional gene product is not produced. It is preferred that if an insertion disrupts the nucleic acid of interest, such a mutation is tolerated in heterozygous form, although it is possible that in some cases homozygosity regarding the fusion gene in which inactivation of the endogenous nucleic acid may alternatively be tolerated. However, as described in the Examples herein, the fusion need not inactivate the nucleic acid of interest.

In an alternative embodiment a fusion gene is created which produces a protein fusion between the product of the reporter sequence and the gene product of the nucleic acid sequence of interest. However, it is preferred that there is an independent start site for translation of the reporter sequence. An additional preferred embodiment is another type of fusion gene called a "transcriptional" fusion, which places an intact reporter gene downstream from the transcriptional start signal of the target gene. The reporter gene fusion genes described herein are preferably quantitated to reflect quantitation of the expression of the nucleic acid of interest. In accordance with an object of the present invention, the system wherein a fusion gene is created between a reporter sequence and endogenous nucleic acid of interest in mammalian systems is utilized to perform high throughput drug screening and/or high throughput genetic screening using cells such as mouse cells in tissue culture, human cells in tissue culture, or intact animals such as mice. In a specific embodiment the drug screening assays for upregulators or downregulators of expression of the reporter sequence, and thus the nucleic acid sequence of interest.

The skilled artisan will readily recognize that the specific nucleotide sequence which is being used to look for upregulators or downregulators does not need to derive from the nucleotide sequence naturally found in the test organism. For example, a human sequence of interest, the sequence coding for apolipoprotein A-I, could be inserted into a mouse, another species or a cell line or cell free extract and then tested in that species or cell line or

cell free extract. Similarly the nucleotide sequence could be from the same species, for example mouse growth hormone tested in mouse for upregulation or downregulation.

Similarly to the specific nucleotide sequence the reporter sequence does not have to be from the same organism, although it can be. For example, as seen in the Examples provided herein, the agouti sequence is used in mice and is a natural mouse sequence. It provides an easy method to select mice, i.e. on coat color. On the other hand, luciferase or green fluorescent protein could be used in a mouse as reporter sequence even though they are not native mouse sequences. Another example would be the use of the human α-fetoprotein or human growth hormone as a reporter sequence in a mouse. When two or more reporter sequences are used the reporter sequences themselves could be from the same species or different species. In addition all, some or none of the reporter sequences could be from the test organism.

One skilled in the art regularly recognizes that test organisms should be selected based on availability in the laboratory as well as the ease of which a marker or reporter sequence can be monitored. In specific embodiments the nucleotide sequence of the present invention is selected from the group of organisms consisting of mammals, non-mammalian animals, plants, insects, aquatic organisms and avian species.

Test organisms can be whole animals (mice and rats), tissue or organs (heart, lung, kidney), tissue slices, cell lines or cell extracts from various organisms. The cell lines and cell free extracts can be from an organism of interest, for example humans, or from a specific tissue or organ of interest such as brain, liver, heart. One cell line that has been useful in the present invention is human stem cells. In many instances cell lines from insects and plants can be used in order to look at regulation of expression of the nucleotide sequence. In many cases the use of organisms which do not have any commonality with the reporter sequence makes the reporter sequence easier to detect.

Another application of the present invention is the use of the system to detect upregulators or downregulators to enhance food production and nutritional value in animals used for food. An example of this would be to search for upregulators of the growth hormone. In this case, once the upregulator is found it could be introduced into the organism of interest, for example sheep, lamb, cow, pig, fish, or shrimp in order to enhance

protein production. Thus, the skilled artisan recognizes there are various applications for the food industry and for animals used as food sources.

Another specific embodiment is the method of treatment of a disease in which there is deficiency of expression or excess expression. The skilled artisan recognizes that a variety of diseases can be treated this way. Some specific examples include: (a) Duchenne muscular dystrophy wherein the utrophin nucleotide sequence can be used to screen for upregulators and then the upregulators can be delivered to patients in pharmacologically effective doses to treat the disease; (b) Alzheimer's whereas the amyloid A beta precursor protein (APP) or presinilin-1 nucleotide sequence can be used to screen for downregulators and the downregulators can then be delivered to patients in pharmacologically effective doses; (c) the apolipoprotein A-I and the LDL receptor can be used to test for compounds useful in the treatment of high cholesterol by using pharmacologically effective doses of upregulators to regulate the production of apolipoprotein A-I or LDL receptor nucleotide sequences; (d) tumor suppressor genes can be upregulated to treat and suppress cancer; and (e) growth hormone disease can be treated using either upregulators or downregulators to treat the various conditions associated with excess or deficient expression of growth hormone.

One skilled in the art recognizes that there are a variety of reporter sequences which may be utilized. One strength of the invention is the use of reporter sequences that are easily detected or quantitated. As in the Examples, this can involve easily visible markers, such as coat color (agouti gene). A variety of other strategies to affect coat color are also possible. These would include the use of secreted proteins such as α-melanocyte-stimulating-hormone (α-MSH) which can affect skin, eye and hair pigmentation when secreted into the circulation from any part of the body. Other reporter genes that are useful include secreted proteins, such as human growth hormone (HGH), human α- fetoprotein (hAFP), or mouse α- fetoprotein (mAFP), and in the mouse where sensitive assays can detect very low levels of compounds in biological fluids or tissues including the blood or urine that may have been secreted by a small number of cells in distant parts of the organism. In this case, the ease of detection does require sampling of biological fluids or tissues including blood or urine from each individual animal for quantitation of the reporter protein or a metabolite secondarily affected by the reporter

protein. Other reporter proteins of interest would obviously include histological markers such as chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), enhanced GFP, blue fluorescent protein, β- galactosidase and β-glucuronidase. Luciferase is another widely utilized reporter gene that is being tested in this system. The reporter gene containing an epitope tag can also be monitored. Examples of epitope tags include HA, myc and Flag.

In general, detection of the reporter means direct measurement or any other effect caused secondarily by the reporter, such as (1) change in the level of a metabolite in blood or urine, or (2) induction of drug resistance or sensitivity in whole animals or tissue culture cells.

The skilled artisan also readily recognizes that the reporter sequence can be inserted anywhere within the specific nucleotide sequence as long as it is expressed by the promoter controlling the specific nucleotide sequence. In specific embodiments it has been found that the reporter sequence can be inserted at either the 5′ or 3′ portion of the coding sequence, in an intron or within the 3′ untranslated portion of the nucleotide sequence. In specific embodiments the reporter sequence is inserted in that part of the sequence which codes for the N-terminal or C-terminal portion of the protein. In other embodiments it is important that the reporter be inserted in a location so that a knockin/fusion product between the knockin (reporter sequence) and the specific nucleotide sequence is formed. This is very helpful in terms of using this method for detecting posttranscriptional events at either the RNA or protein level. One skilled in the art also readily recognizes that when the receptor sequence becomes a knockin (i.e. inserted into the specific nucleotide sequence), the specific nucleotide sequence itself may not be expressed. As long as this is not lethal to the organism being tested, the methods of the present invention will work to allow for the termination of upregulators and downregulators of the promoter which regulates the specific sequence. One skilled in the art recognizes that if it is lethal then the location for insertion needs to be adjusted. Examples of this are shown below.

It is readily apparent to one skilled in the art that the relationship between a promoter and its regulatory elements and the reporter sequence may be addressed by using methods other than homologous recombination in ES cells or other cells to prepare the knockin/fusion sequence. It is obvious that any other method of recombination that would

test the activity of the promoter or test for posttranscriptional effects by linking a reporter sequence to the promoter, to the transcript, or to the protein would be included within the scope of the present invention. Also, the invention may be utilized *in vitro* with fragments of genes or cell extracts.

The skilled artisan recognizes there are a variety of ways to practice the present invention. Some examples include introducing reporter sequence into ES cells by homologous or other recombination and obtain germline transmission. Then a skilled artisan may study animals and/or cultured cells derived from the animals. Alternatively the reporter sequence can be introduced into cultured cells either by homologous or non-homologous recombination or as an extrachromosal nucleic acid or similar strategy and the cultured cells studied. This can be done for animals, including humans, and especially human stem cells.

In the transcriptional application, the promoter or other regulatory element such as an enhancer will be used to control the reporter sequence. This may involve a fusion transcript or not. Some variations include: (1) the promoter of the specific nucleotide sequence of interest is fused to reporter sequence such that ATG codon for start of translation is part of the reporter sequence; (2) similar to (1) except that the ATG codon of the specific nucleotide sequence of interest is used resulting in a small but functionally insignificant N-terminal addition of amino acids to the reporter product; and (3) an enhancer or silencer from the specific nucleotide sequence of interest is fused to a minimal promoter to express the reporter sequence such that the promoter of the specific nucleotide sequence of interest is not used while testing other *cis* elements such as enhancers or silencers.

One embodiment of this invention is to carry out mutagenesis in mice, other mammalian organisms, and other species. The Examples provided herein present ENU mutagenesis in the mouse, which is well known by skilled artisans to be one of the most powerful methods of introducing mutations into the mouse at the present time. However, it is also known by the skilled artisan that it is very difficult to identify, recover and demonstrate the cause and effect relationship with point mutations introduced by ENU mutagenesis. These mutations are readily mapped and are increasingly easily characterized with progress on the sequencing of the mouse genome, but other mutagenesis strategies have potential advantages. Insertional mutagenesis, sometimes referred to as transposon tagging,

has the major advantage of providing for the easy recovery of the mutated gene. Thus, a wide variety of mutagens, including chemicals, X-ray, UV, and biological agents are useful. One skilled in the art recognizes that any mutagenizing method wherein the mutant sequence can be identified is useful in this procedure.

NUCLEIC ACID-BASED EXPRESSION SYSTEMS

1. Vectors

The term "vector" is used to refer to a carrier nucleic acid molecule into which a nucleic acid sequence can be inserted for introduction into a cell where it can be replicated. A nucleic acid sequence can be "exogenous," which means that it is foreign to the cell into which the vector is being introduced or that the sequence is homologous to a sequence in the cell but in a position within the host cell nucleic acid in which the sequence is ordinarily not found. Vectors include plasmids, cosmids, viruses (bacteriophage, animal viruses, and plant viruses), and artificial chromosomes (e.g., YACs). One of skill in the art would be well equipped to construct a vector through standard recombinant techniques, which are described in Maniatis et al., 1988 and Ausubel et al., 1994, both incorporated herein by reference.

The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are then translated into a protein, polypeptide, or peptide. In other cases, these sequences are not translated, for example, in the production of antisense molecules or ribozymes. Expression vectors can contain a variety of "control sequences," which refer to nucleic acid sequences necessary for the transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well and are described *infra*.

a. Promoters and Enhancers

A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other

transcription factors. The phrases "operatively positioned," "operatively linked," "under control," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. A promoter may or may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence.

A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment and/or exon. Such a promoter can be referred to as "endogenous." Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell, and promoters or enhancers not "naturally occurring," i.e., containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent 4,683,202, U.S. Patent 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated the control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

Naturally, it will be important to employ a promoter and/or enhancer that effectively directs the expression of the DNA segment in the cell type, organelle, and organism chosen for expression. Those of skill in the art of molecular biology generally know the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook et al. (1989), incorporated herein by reference. The promoters employed may be

constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

The identity of tissue-specific promoters or elements, as well as assays to characterize their activity, is well known to those of skill in the art. Examples of such regions include the human LIMK2 gene (Nomoto et al. 1999), the somatostatin receptor 2 gene (Kraus et al., 1998), murine epididymal retinoic acid-binding gene (Lareyre et al., 1999), human CD4 (Zhao-Emonet et al., 1998), mouse alpha2 (XI) collagen (Tsumaki, et al., 1998), D1A dopamine receptor gene (Lee, et al., 1997), insulin-like growth factor II (Wu et al., 1997), human platelet endothelial cell adhesion molecule-1 (Almendro et al., 1996).

b. Initiation Signals and Internal Ribosome Binding Sites

A specific initiation signal also may be required for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

In certain embodiments of the invention, the use of internal ribosome entry sites (IRES) elements are used to create multigene, or polycistronic, messages. IRES elements are able to bypass the ribosome scanning model of 5′ methylated Cap dependent translation and begin translation at internal sites (Pelletier and Sonenberg, 1988). IRES elements from two members of the picornavirus family (polio and encephalomyocarditis) have been described (Pelletier and Sonenberg, 1988), as well an IRES from a mammalian message (Macejak and Sarnow, 1991). IRES elements can be linked to heterologous open reading frames. Multiple open reading frames can be transcribed together, each separated by an

IRES, creating polycistronic messages. By virtue of the IRES element, each open reading frame is accessible to ribosomes for efficient translation. Multiple genes can be efficiently expressed using a single promoter/enhancer to transcribe a single message (see U.S. Patent 5,925,565 and 5,935,819, herein incorporated by reference).

c. Multiple Cloning Sites

Vectors can include a multiple cloning site (MCS), which is a nucleic acid region that contains multiple restriction enzyme sites, any of which can be used in conjunction with standard recombinant technology to digest the vector. (See Carbonelli et al., 1999, Levenson et al., 1998, and Cocea, 1997, incorporated herein by reference.) "Restriction enzyme digestion" refers to catalytic cleavage of a nucleic acid molecule with an enzyme that functions only at specific locations in a nucleic acid molecule. Many of these restriction enzymes are commercially available. Use of such enzymes is widely understood by those of skill in the art. Frequently, a vector is linearized or fragmented using a restriction enzyme that cuts within the MCS to enable exogenous sequences to be ligated to the vector. "Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments, which may or may not be contiguous with each other. Techniques involving restriction enzymes and ligation reactions are well known to those of skill in the art of recombinant technology.

d. Splicing Sites

Most transcribed eukaryotic RNA molecules will undergo RNA splicing to remove introns from the primary transcripts. A skilled artisan is aware that genomic eukaryotic sequences may require donor and/or acceptor splicing sites to ensure proper processing of the transcript for protein expression. (See Chandler et al., 1997, herein incorporated by reference.)

e. Polyadenylation Signals

In expression, one will typically include a polyadenylation signal to effect proper polyadenylation of the transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and/or any such sequence may be

employed. Preferred embodiments include the SV40 polyadenylation signal and/or the bovine growth hormone polyadenylation signal, convenient and/or known to function well in various target cells. Also contemplated as an element of the expression cassette is a transcriptional termination site. These elements can serve to enhance message levels and/or to minimize read through from the cassette into other sequences.

f. Origins of Replication

In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated. Alternatively an autonomously replicating sequence (ARS) can be employed if the host cell is yeast.

g. Selectable and Screenable Markers

In certain embodiments of the invention, the cells contain nucleic acid construct of the present invention, a cell may be identified *in vitro* or *in vivo* by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers. In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes such as herpes simplex virus thymidine kinase (tk) or chloramphenicol acetyltransferase (CAT) may be utilized. One of skill in the art would also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed

simultaneously with the nucleic acid encoding a gene product. Further examples of selectable and screenable markers are well known to one of skill in the art.

2. Host Cells

As used herein, the terms "cell," "cell line," and "cell culture" may be used interchangeably. All of these term also include their progeny, which is any and all subsequent generations. It is understood that all progeny may not be identical due to deliberate or inadvertent mutations. In the context of expressing a heterologous nucleic acid sequence, "host cell" refers to a prokaryotic or eukaryotic cell, and it includes any transformable organisms that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell can, and has been, used as a recipient for vectors. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

Host cells may be derived from prokaryotes or eukaryotes, depending upon whether the desired result is replication of the vector or expression of part or all of the vector-encoded nucleic acid sequences. Numerous cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials (www.atcc.org). An appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5a, JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE® Competent Cells and SOLOPACKä Gold Cells (STRATAGENE®, La Jolla). Alternatively, bacterial cells such as *E. coli* LE392 could be used as host cells for phage viruses.

Examples of eukaryotic host cells for replication and/or expression of a vector include HeLa, NIH3T3, Jurkat, 293, Cos, CHO, Saos, and PC12. Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector may be used in conjunction with either a eukaryotic or prokaryotic host cell, particularly one that is permissive for replication or expression of the vector.

Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

3. Expression Systems

Numerous expression systems exist that comprise at least a part or all of the compositions discussed above. Prokaryote- and/or eukaryote-based systems can be employed for use with the present invention to produce nucleic acid sequences, or their cognate polypeptides, proteins and peptides. Many such systems are commercially and widely available.

The insect cell/baculovirus system can produce a high level of protein expression of a heterologous nucleic acid segment, such as described in U.S. Patent No. 5,871,986, 4,879,236, both herein incorporated by reference, and which can be bought, for example, under the name MAXBAC® 2.0 from INVITROGEN® and BACPACK™ BACULOVIRUS EXPRESSION SYSTEM FROM CLONTECH®.

Other examples of expression systems include STRATAGENE®'s COMPLETE CONTROLä Inducible Mammalian Expression System, which involves a synthetic ecdysone-inducible receptor, or its pET Expression System, an E. coli expression system. Another example of an inducible expression system is available from INVITROGEN®, which carries the T-REX™ (tetracycline-regulated expression) System, an inducible mammalian expression system that uses the full-length CMV promoter. INVITROGEN® also provides a yeast expression system called the Pichia methanolica Expression System, which is designed for high-level production of recombinant proteins in the methylotrophic yeast *Pichia methanolica*. One of skill in the art would know how to express a vector, such as an expression construct, to produce a nucleic acid sequence or its cognate polypeptide, protein, or peptide.

Derivatives of Promoter Sequences

One aspect of the invention provides derivatives of specific promoters. One means for preparing derivatives of such promoters comprises introducing mutations into the promoter sequences. Such mutants may potentially have enhanced, reduced, or altered function relative to the native sequence, or alternatively, may be silent with regard to function.

Mutagenesis may be carried out at random and the mutagenized sequences screened for function. Alternatively, particular sequences which provide the promoter region with desirable expression characteristics could be identified and these or similar sequences introduced into other related or non-related sequences via mutation. Similarly, non-essential elements may be deleted without significantly altering the function of the promoter. It is further contemplated that one could mutagenize these sequences in order to enhance their utility in expressing transgenes, especially in a gene therapy construct in humans.

The means for mutagenizing a DNA segment comprising a specific promoter sequence are well-known to those of skill in the art. Mutagenesis may be performed in accordance with any of the techniques known in the art, such as, and not limited to, synthesizing an oligonucleotide having one or more mutations within the sequence of a particular regulatory region. In particular, site-specific mutagenesis is a technique useful in the preparation of promoter mutants, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, by introducing one or more nucleotide sequence changes into the DNA.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 13 to about 75 nucleotides or more in length is preferred, with about 10 to about 25 or more residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form.

Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids also are routinely employed in site directed mutagenesis to eliminate the step of transferring the gene of interest from a plasmid to a phage.

Alternatively, the use of PCR™ with commercially available thermostable enzymes such as Taq polymerase may be used to incorporate a mutagenic oligonucleotide primer into an amplified DNA fragment that can then be cloned into an appropriate cloning or expression vector. The PCR™-mediated mutagenesis procedures of Tomic *et al.* (1990) and Upender *et al.* (1995) provide two examples of such protocols.

The preparation of sequence variants of the selected promoter or intron-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of DNA sequences may be obtained. For example, recombinant vectors encoding the desired promoter sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Patent No. 4,237,224, incorporated herein by reference. A number of template dependent processes are available to amplify the target sequences of interest present in a sample, such methods being well known in the art and specifically disclosed herein.

One efficient, targeted means for preparing mutagenized promoters or enhancers relies upon the identification of putative regulatory elements within the target sequence. These can be identified, for example, by comparison with known promoter sequences. Sequences which are shared among genes with similar functions or expression patterns are likely candidates for the binding of transcription factors and are likely elements to confer tissue specific expression patterns.

One of skill in the art will recognize that regulatory elements may be included in regions of the gene other than the 5'-untranslated region, and comparison of coding and

3'-noncoding regions of genes may identify putative regulatory elements. Confirmation of putative regulatory elements can be achieved by deletion, duplication, or other alteration or mutation of each putative regulatory region followed by functional analysis of each construct by assay of a reporter gene which is functionally attached to each construct. As such, once a starting promoter sequence is provided, any of a number of different functional mutants of the starting sequence could be readily prepared using methods well known in the art (Zhang *et al*, 1997).

Mutation, alteration, duplication, or truncation mutants the promoter region of the invention could be randomly prepared or prepared by selection of regions identified as containing putative regulatory elements, and then assayed. With this strategy, a series of constructs are prepared, each containing a different portion of the clone (a subclone), either mutated or altered or containing wild-type sequence, and these constructs are then screened for activity. A suitable means for screening for activity is to attach such a promoter construct to a selectable or screenable marker gene and to assay for gene expression.

Other assays may be used to identify responsive elements in a promoter region or gene. Such assays will be known to those of skill in the art (see for example, Sambrook *et al.*, 1989; Zhang *et al*, 1997; Shan *et al.*, 1997; Dai and Burnstein, 1996; Cleutjens *et al.*, 1997; Ng *et al.*, 1994; Shida *et al.*, 1993), and include DNase I footprinting studies, Elecromobility Shift Assay patterns (EMSA), the binding pattern of purified transcription factors, effects of specific transcription factor antibodies in inhibiting the binding of a transcription factor to a putative responsive element, Western analysis, nuclear run-on assays, and DNA methylation interference analysis.

Preferred promoter constructs may be identified that retain the desired, or even enhanced, activity (e.g., tissue-specific expression). The smallest segment required for activity may be identified through comparison of the selected deletion or mutation constructs. Once identified, such segments may be duplicated, mutated, or combined with other known or regulatory elements and assayed for activity or regulatory properties. Promoter region sequences used to identify regulatory elements can also be used to identify and isolate transcription factors that bind a putative regulatory sequence or element, according to standard methods of protein purification, such as affinity chromatography, as discussed above.

Preferably, identified promoter region sequences, whether used alone or combined with additional promoters, enhancers, or regulatory elements, will be induced and/or regulated by an external agent, such as a hormone, transcription factor, enzyme, or pharmaceutical agent, to express operatively linked genes or sequences (Zhang *et al.*, 1997; Shan *et al.*, 1997). Alternatively, such a construct may be designed to cease expression upon exposure to an external agent.

Additionally, deletion mutants may be produced and assayed essentially according to Matusik (U.S. Patent No. 5,783,681, July 21, 1998). Plasmids may be constructed containing the promoter adjacent to a reporter gene, for example CAT. The construct may be designed to contain additional regulatory sequences, such as polyadenylation, termination and cleavage signals. Deletion mutants may be prepared by a time course treatment of the isolated prostate specific transglutaminase promoter with Bal 31 exonuclease (for 15, 30, 45, 60 and 75 seconds, for example). Following limited digestion, the promoter sequence may be ligated to appropriate linker sequences and reinserted into the CAT expression vector. After transformation into an appropriate host cell, such as *E. coli*, clones containing deletion mutants may be screened and their plasmid DNAs digested with restriction enzymes that cut at sites flanking the promoter sequence. Promoter size may be determined by agarose gel electrophoresis according to standard techniques.

Following selection of a range of deletion mutants of varying size, the activities of the deleted promoters for expression of the linked CAT gene may be determined according to standard protocols.

The precise nature of the deleted portion of the promoter may be determined using standard DNA sequencing, such as Sanger dideoxy termination sequencing, to identify which promoter sequences have been removed in each of the assayed deletion mutants. Thus, a correlation may be obtained between the presence or absence of specific elements within the promoter sequence and changes in activity of the linked reporter gene.

Assays of Gene Expression

Assays may be employed within the scope of the instant invention for determination of the relative efficiency of gene expression. For example, assays may be used to determine the efficacy of deletion mutants of specific promoter regions in directing expression of

operably linked genes. Similarly, one could produce random or site-specific mutants of promoter regions and assay the efficacy of the mutants in the expression of an operably linked gene. Alternatively, assays could be used to determine the function of a promoter region in enhancing gene expression when used in conjunction with various different regulatory elements, enhancers, and exogenous genes.

Gene expression may be determined by measuring the production of RNA, protein or both. The gene product (RNA or protein) may be isolated and/or detected by methods well known in the art. Following detection, one may compare the results seen in a given cell line or individual with a statistically significant reference group of non-transformed control cells. Alternatively, one may compare production of RNA or protein products in cell lines transformed with the same gene operably linked to various mutants of a promoter sequence. In this way, it is possible to identify regulatory regions within a novel promoter sequence by their effect on the expression of an operably linked gene.

In certain embodiments, it will be desirable to use genes whose expression is naturally linked to a given promoter or other regulatory element. For example, a prostate specific promoter may be operably linked to a gene that is normally expressed in prostate tissues. Alternatively, marker genes may be used for assaying promoter activity. Using, for example, a selectable marker gene, one could quantitatively determine the resistance conferred upon a tissue culture cell line or animal cell by a construct comprising the selectable marker gene operably linked to the promoter to be assayed. Alternatively, various tissue culture cell line or animal parts could be exposed to a selective agent and the relative resistance provided in these parts quantified, thereby providing an estimate of the tissue specific expression of the promoter.

Screenable markers constitute another efficient means for quantifying the expression of a given gene. Potentially any screenable marker could be expressed and the marker gene product quantified, thereby providing an estimate of the efficiency with which the promoter directs expression of the gene. Quantification can readily be carried out using either visual means, or, for example, a photon counting device.

A preferred screenable marker gene for use with the current invention is β-glucuronidase (GUS). Detection of GUS activity can be performed histochemically using 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) as the substrate for the GUS enzyme,

yielding a blue precipitate inside of cells containing GUS activity. This assay has been described in detail (Jefferson, 1987). The blue coloration can then be visually scored, and estimates of expression efficiency thereby provided. GUS activity also can be determined by immunoblot analysis or a fluorimetric GUS specific activity assay (Jefferson, 1987).

Nucleic Acid Detection

In accordance with the object of the present invention, there is an embodiment wherein the expression of a reporter sequence, an endogenous nucleic acid sequence of interest, or a fusion between all of parts of the two are detected.

1. Hybridization

The use of a probe or primer of between 13 and 100 nucleotides, preferably between 17 and 100 nucleotides in length, or in some aspects of the invention up to 1-2 kilobases or more in length, allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over contiguous stretches greater than 20 bases in length are generally preferred, to increase stability and/or selectivity of the hybrid molecules obtained. One will generally prefer to design nucleic acid molecules for hybridization having one or more complementary sequences of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs or to provide primers for amplification of DNA or RNA from samples. Depending on the application envisioned, one would desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of the probe or primers for the target sequence.

For applications requiring high selectivity, one will typically desire to employ relatively high stringency conditions to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Such high stringency conditions tolerate little, if any, mismatch between the probe or primers and the template or target strand and would

be particularly suitable for isolating specific genes or for detecting specific mRNA transcripts. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide.

For certain applications, for example, site-directed mutagenesis, it is appreciated that lower stringency conditions are preferred. Under these conditions, hybridization may occur even though the sequences of the hybridizing strands are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and/or decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Hybridization conditions can be readily manipulated depending on the desired results.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl2, 1.0 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, at temperatures ranging from approximately 40°C to about 72°C.

In certain embodiments, it will be advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples.

In general, it is envisioned that the probes or primers described herein will be useful as reagents in solution hybridization, as in PCR™, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a

selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The conditions selected will depend on the particular circumstances (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, etc.). Optimization of hybridization conditions for the particular application of interest is well known to those of skill in the art. After washing of the hybridized molecules to remove non-specifically bound probe molecules, hybridization is detected, and/or quantified, by determining the amount of bound label. Representative solid phase hybridization methods are disclosed in U.S. Patent Nos. 5,843,663, 5,900,481 and 5,919,626. Other methods of hybridization that may be used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,849,481, 5,849,486 and 5,851,772. The relevant portions of these and other references identified in this section of the Specification are incorporated herein by reference.

2. Amplification of Nucleic Acids

Nucleic acids used as a template for amplification may be isolated from cells, tissues or other samples according to standard methodologies (Sambrook *et al.*, 1989). In certain embodiments, analysis is performed on whole cell or tissue homogenates or biological fluid samples without substantial purification of the template nucleic acid. The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to first convert the RNA to a complementary DNA.

The term "primer," as used herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty and/or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded and/or single-stranded form, although the single-stranded form is preferred.

Pairs of primers designed to selectively hybridize to a nucleic acid of interest are contacted with the template nucleic acid under conditions that permit selective hybridization. Depending upon the desired application, high stringency hybridization conditions may be selected that will only allow hybridization to sequences that are completely complementary to the primers. In other embodiments, hybridization may occur under reduced stringency to allow for amplification of nucleic acids contain one or more mismatches with the primer

sequences. Once hybridized, the template-primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

The amplification product may be detected or quantified. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical and/or thermal impulse signals (Affymax technology; Bellus, 1994).

A number of template dependent processes are available to amplify the oligonucleotide sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1990, each of which is incorporated herein by reference in their entirety.

A reverse transcriptase PCR™ amplification procedure may be performed to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable DNA polymerases. These methods are described in WO 90/07641. Polymerase chain reaction methodologies are well known in the art. Representative methods of RT-PCR are described in U.S. Patent No. 5,882,864.

Another method for amplification is ligase chain reaction ("LCR"), disclosed in European Application No. 320 308, incorporated herein by reference in its entirety. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence. A method based on PCR™ and oligonucleotide ligase assay (OLA), disclosed in U.S. Patent 5,912,148, may also be used.

Alternative methods for amplification of target nucleic acid sequences that may be used in the practice of the present invention are disclosed in U.S. Patent Nos. 5,843,650, 5,846,709, 5,846,783, 5,849,546, 5,849,497, 5,849,547, 5,858,652, 5,866,366, 5,916,776, 5,922,574, 5,928,905, 5,928,906, 5,932,451, 5,935,825, 5,939,291 and 5,942,391, GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety.

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, may also be used as an amplification method in the present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence which may then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention (Walker et al., 1992). Strand Displacement Amplification (SDA), disclosed in U.S. Patent No. 5,916,779, is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, i.e., nick translation.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh *et al.*, 1989; Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference in their entirety). Davey *et al.*, European Application No. 329 822 disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention.

Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter region/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, i.e., new templates are not produced from the resultant RNA transcripts. Other amplification methods include "race" and "one-sided PCR" (Frohman, 1990; Ohara et al., 1989).

3. Detection of Nucleic Acids

Following any amplification, it may be desirable to separate the amplification product from the template and/or the excess primer. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis

using standard methods (Sambrook *et al.*, 1989). Separated amplification products may be cut out and eluted from the gel for further manipulation. Using low melting point agarose gels, the separated band may be removed by heating the gel, followed by extraction of the nucleic acid.

Separation of nucleic acids may also be effected by chromatographic techniques known in art. There are many kinds of chromatography which may be used in the practice of the present invention, including adsorption, partition, ion-exchange, hydroxylapatite, molecular sieve, reverse-phase, column, paper, thin-layer, and gas chromatography as well as HPLC.

In certain embodiments, the amplification products are visualized. A typical visualization method involves staining of a gel with ethidium bromide and visualization of bands under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the separated amplification products can be exposed to x-ray film or visualized under the appropriate excitatory spectra.

In one embodiment, following separation of amplification products, a labeled nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, or another binding partner carrying a detectable moiety.

In particular embodiments, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art. See Sambrook *et al.*, 1989. One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

Other methods of nucleic acid detection that may be used in the practice of the instant invention are disclosed in U.S. Patent Nos. 5,840,873, 5,843,640, 5,843,651, 5,846,708, 5,846,717, 5,846,726, 5,846,729, 5,849,487, 5,853,990, 5,853,992, 5,853,993, 5,856,092, 5,861,244, 5,863,732, 5,863,753, 5,866,331, 5,905,024, 5,910,407, 5,912,124, 5,912,145,

5,919,630, 5,925,517, 5,928,862, 5,928,869, 5,929,227, 5,932,413 and 5,935,791, each of which is incorporated herein by reference.

4. Other Assays

Other methods for genetic screening may be used within the scope of the present invention, for example, to detect mutations in genomic DNA, cDNA and/or RNA samples. Methods used to detect point mutations include denaturing gradient gel electrophoresis ("DGGE"), restriction fragment length polymorphism analysis ("RFLP"), chemical or enzymatic cleavage methods, direct sequencing of target regions amplified by PCRTM (see above), single-strand conformation polymorphism analysis ("SSCP") and other methods well known in the art.

One method of screening for point mutations is based on RNase cleavage of base pair mismatches in RNA/DNA or RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single or multiple base point mutations.

U.S. Patent No. 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as positive.

Other investigators have described the use of RNase I in mismatch assays. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is reported to cleave three out of four known mismatches. Others have described using the MutS protein or other DNA-repair enzymes for detection of single-base mismatches.

Alternative methods for detection of deletion, insertion or substititution mutations that may be used in the practice of the present invention are disclosed in U.S. Patent Nos.

5,849,483, 5,851,770, 5,866,337, 5,925,525 and 5,928,870, each of which is incorporated herein by reference in its entirety.

5. Kits

All the essential materials and/or reagents required for detecting a nucleic acid sequence of interest in a sample may be assembled together in a kit. This generally will comprise a probe or primers designed to hybridize specifically to individual nucleic acids of interest or reporter sequence in the practice of the present invention. Also included may be enzymes suitable for amplifying nucleic acids, including various polymerases (reverse transcriptase, Taq, etc.), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification. Such kits may also include enzymes and other reagents suitable for detection of specific nucleic acids or amplification products. Such kits generally will comprise, in suitable means, distinct containers for each individual reagent or enzyme as well as for each probe or primer pair.

Mutagenesis

Where employed, mutagenesis will be accomplished by a variety of standard, mutagenic procedures. Mutation is the process whereby changes occur in the quantity or structure of an organism. Mutation can involve modification of the nucleotide sequence of a single gene, blocks of genes or whole chromosome. Changes in single genes may be the consequence of point mutations which involve the removal, addition or substitution of a single nucleotide base within a DNA sequence, or they may be the consequence of changes involving the insertion or deletion of large numbers of nucleotides.

Mutations can arise spontaneously as a result of events such as errors in the fidelity of DNA replication or the movement of transposable genetic elements (transposons) within the genome. They also are induced following exposure to chemical or physical mutagens. Such mutation-inducing agents include ionizing radiations, ultraviolet light and a diverse array of chemical such as alkylating agents and polycyclic aromatic hydrocarbons, all of which are capable of interacting either directly or indirectly (generally following some metabolic biotransformations) with nucleic acids. The DNA lesions induced by such environmental agents may lead to modifications of base sequence when the affected DNA

is replicated or repaired and thus to a mutation. Mutation also can be site-directed through the use of particular targeting methods.

A. Random Mutagenesis

i) Insertional Mutagenesis

Insertional mutagenesis is based on the inactivation of a gene via insertion of a known DNA fragment. Because it involves the insertion of some type of DNA fragment, the mutations generated are generally loss-of-function, rather than gain-of-function mutations. However, there are several examples of insertions generating gain-of-function mutations (Oppenheimer et al. 1991). Insertion mutagenesis has been very successful in bacteria and *Drosophila* (Cooley et al. 1988) and recently has become a powerful tool in corn (Schmidt et al. 1987); Arabidopsis; (Marks et al., 1991; Koncz et al. 1990); and Antirrhinum (Sommer et al. 1990).

Transposable genetic elements are DNA sequences that can move (transpose) from one place to another in the genome of a cell. The first transposable elements to be recognized were the Activator/Dissociation elements of Zea mays (McClintock, 1957). Since then, they have been identified in a wide range of organisms, both prokaryotic and eukaryotic.

Transposable elements in the genome are characterized by being flanked by direct repeats of a short sequence of DNA that has been duplicated during transposition and is called a target site duplication. Virtually all transposable elements whatever their type, and mechanism of transposition, make such duplications at the site of their insertion. In some cases the number of bases duplicated is constant, in other cases it may vary with each transposition event. Most transposable elements have inverted repeat sequences at their termini. these terminal inverted repeats may be anything from a few bases to a few hundred bases long and in many cases they are known to be necessary for transposition.

Prokaryotic transposable elements have been most studied in *E. coli* and Gram negative bacteria, but also are present in Gram positive bacteria. They are generally termed insertion sequences if they are less than about 2 kB long, or transposons if they are longer. Bacteriophages such as mu and D108, which replicate by transposition, make up a third type of transposable element. elements of each type encode at least one polypeptide a

transposase, required for their own transposition. Transposons often further include genes coding for function unrelated to transposition, for example, antibiotic resistance genes.

Transposons can be divided into two classes according to their structure. First, compound or composite transposons have copies of an insertion sequence element at each end, usually in an inverted orientation. These transposons require transposases encoded by one of their terminal IS elements. The second class of transposon have terminal repeats of about 30 base pairs and do not contain sequences from IS elements.

Transposition usually is either conservative or replicative, although in some cases it can be both. In replicative transposition, one copy of the transposing element remains at the donor site, and another is inserted at the target site. In conservative transposition, the transposing element is excised from one site and inserted at another. Eukaryotic elements also can be classified according to their structure and mechanism of transportation. The primary distinction is between elements that transpose via an RNA intermediate, and elements that transpose directly from DNA to DNA.

Elements that transpose via an RNA intermediate often are referred to as retrotransposons, and their most characteristic feature is that they encode polypeptides that are believed to have reverse transcriptionase activity. There are two types of retrotransposon. Some resemble the integrated proviral DNA of a retrovirus in that they have long direct repeat sequences, long terminal repeats (LTRs), at each end. The similarity between these retrotransposons and proviruses extends to their coding capacity. They contain sequences related to the gag and pol genes of a retrovirus, suggesting that they transpose by a mechanism related to a retroviral life cycle. Retrotransposons of the second type have no terminal repeats. They also code for gag- and pol-like polypeptides and transpose by reverse transcription of RNA intermediates, but do so by a mechanism that differs from that or retrovirus-like elements. Transposition by reverse transcription is a replicative process and does not require excision of an element from a donor site.

Transposable elements are an important source of spontaneous mutations, and have influenced the ways in which genes and genomes have evolved. They can inactivate genes by inserting within them, and can cause gross chromosomal rearrangements either directly, through the activity of their transposases, or indirectly, as a result of recombination between copies of an element scattered around the genome. Transposable elements that excise often

do so imprecisely and may produce alleles coding for altered gene products if the number of bases added or deleted is a multiple of three.

Transposable elements themselves may evolve in unusual ways. If they were inherited like other DNA sequences, then copies of an element in one species would be more like copies in closely related species than copies in more distant species. This is not always the case, suggesting that transposable elements are occasionally transmitted horizontally from one species to another.

ii) Chemical mutagenesis

Chemical mutagenesis offers certain advantages, such as the ability to find a full range of mutant alleles with degrees of phenotypic severity, and is facile and inexpensive to perform. The majority of chemical carcinogens produce mutations in DNA. Benzo[a]pyrene, N-acetoxy-2-acetyl aminofluorene and aflotoxin B1 cause GC to TA transversions in bacteria and mammalian cells. Benzo[a]pyrene also can produce base substitutions such as AT to TA. N-nitroso compounds produce GC to AT transitions. Alkylation of the O4 position of thymine induced by exposure to n-nitrosoureas results in TA to CG transitions.

A high correlation between mutagenicity and carcinogenity is the underlying assumption behind the Ames test (McCann et al., 1975) which speedily assays for mutants in a bacterial system, together with an added rat liver homogenate, which contains the microsomal cytochrome P450, to provide the metabolic activation of the mutagens where needed.

In vertebrates, several carcinogens have been found to produce mutation in the ras proto-oncogene. N-nitroso-N-methyl urea induces mammary, prostate and other carcinomas in rats with the majority of the tumors showing a G to A transition at the second position in codon 12 of the Ha-ras oncogene. Benzo[a]pyrene-induced skin tumors contain A to T transformation in the second codon of the *Ha-ras* gene.

iii) Radiation Mutagenesis

The integrity of biological molecules is degraded by the ionizing radiation. Adsorption of the incident energy leads to the formation of ions and free radicals, and

breakage of some covalent bonds. Susceptibility to radiation damage appears quite variable between molecules, and between different crystalline forms of the same molecule. It depends on the total accumulated dose, and also on the dose rate (as once free radicals are present, the molecular damage they cause depends on their natural diffusion rate and thus upon real time). Damage is reduced and controlled by making the sample as cold as possible.

Ionizing radiation causes DNA damage and cell killing, generally proportional to the dose rate. Ionizing radiation has been postulated to induce multiple biological effects by direct interaction with DNA, or through the formation of free radical species leading to DNA damage (Hall, 1988). These effects include gene mutations, malignant transformation, and cell killing. Although ionizing radiation has been demonstrated to induce expression of certain DNA repair genes in some prokaryotic and lower eukaryotic cells, little is known about the effects of ionizing radiation on the regulation of mammalian gene expression (Borek, 1985). Several studies have described changes in the pattern of protein synthesis observed after irradiation of mammalian cells. For example, ionizing radiation treatment of human malignant melanoma cells is associated with induction of several unidentified proteins (Boothman et al., 1989). Synthesis of cyclin and co-regulated polypeptides is suppressed by ionizing radiation in rat REF52 cells, but not in oncogene-transformed REF52 cell lines (Lambert and Borek, 1988). Other studies have demonstrated that certain growth factors or cytokines may be involved in x-ray-induced DNA damage. In this regard, platelet-derived growth factor is released from endothelial cells after irradiation (Witte, et al., 1989).

In the present invention, the term "ionizing radiation" means radiation comprising particles or photons that have sufficient energy or can produce sufficient energy via nuclear interactions to produce ionization (gain or loss of electrons). An exemplary and preferred ionizing radiation is an x-radiation. The amount of ionizing radiation needed in a given cell generally depends upon the nature of that cell. Typically, an effective expression-inducing dose is less than a dose of ionizing radiation that causes cell damage or death directly. Means for determining an effective amount of radiation are well known in the art.

In a certain embodiments, an effective expression inducing amount is from about 2 to about 30 Gray (Gy) administered at a rate of from about 0.5 to about 2 Gy/minute. Even

more preferably, an effective expression inducing amount of ionizing radiation is from about 5 to about 15 Gy. In other embodiments, doses of 2-9 Gy are used in single doses. An effective dose of ionizing radiation may be from 10 to 100 Gy, with 15 to 75 Gy being preferred, and 20 to 50 Gy being more preferred.

Any suitable means for delivering radiation to a tissue may be employed in the present invention in addition to external means. For example, radiation may be delivered by first providing a radiolabeled antibody that immunoreacts with an antigen of the tumor, followed by delivering an effective amount of the radiolabeled antibody to the tumor. In addition, radioisotopes may be used to deliver ionizing radiation to a tissue or cell.

iv) *In Vitro* Scanning Mutagenesis

Random mutagenesis also may be introduced using error prone PCR (Cadwell and Joyce, 1992). The rate of mutagenesis may be increased by performing PCR in multiple tubes with dilutions of templates.

One particularly useful mutagenesis technique is alanine scanning mutagenesis in which a number of residues are substituted individually with the amino acid alanine so that the effects of losing side-chain interactions can be determined, while minimizing the risk of large-scale perturbations in protein conformation (Cunningham et al., 1989).

In recent years, techniques for estimating the equilibrium constant for ligand binding using minuscule amounts of protein have been developed (Blackburn *et al.*, 1991; U.S. Patents 5,221,605 and 5,238,808). The ability to perform functional assays with small amounts of material can be exploited to develop highly efficient, in vitro methodologies for the saturation mutagenesis of antibodies. The inventors bypassed cloning steps by combining PCR mutagenesis with coupled *in vitro* transcription/translation for the high throughput generation of protein mutants. Here, the PCR products are used directly as the template for the *in vitro* transcription/translation of the mutant single chain antibodies. Because of the high efficiency with which all 19 amino acid substitutions can be generated and analyzed in this way, it is now possible to perform saturation mutagenesis on numerous residues of interest, a process that can be described as in vitro scanning saturation mutagenesis (Burks *et al.*, 1997).

In vitro scanning saturation mutagenesis provides a rapid method for obtaining a large amount of structure-function information including: (i) identification of residues that modulate ligand binding specificity, (ii) a better understanding of ligand binding based on the identification of those amino acids that retain activity and those that abolish activity at a given location, (iii) an evaluation of the overall plasticity of an active site or protein subdomain, (iv) identification of amino acid substitutions that result in increased binding.

v) Random Mutagenesis by Fragmentation and Reassembly

A method for generating libraries of displayed polypeptides is described in U.S. Patent 5,380,721. The method comprises obtaining polynucleotide library members, pooling and fragmenting the polynucleotides, and reforming fragments therefrom, performing PCR amplification, thereby homologously recombining the fragments to form a shuffled pool of recombined polynucleotides.

B. Site-Directed Mutagenesis

Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of protein-ligand interactions (Wells, 1996, Braisted et al., 1996). The technique provides for the preparattion and testing of sequence variants by introducing one or more nucleotide sequence changes into a selected DNA.

Site-specific mutagenesis uses specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent, unmodified nucleotides. In this way, a primer sequence is provided with sufficient size and complexity to form a stable duplex on both sides of the deletion junction being traversed. A primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded plasmids are also routinely employed in site-directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, one first obtains a single-stranded vector, or melts two strands of a double-stranded vector, which includes within its sequence a DNA sequence encoding the desired protein or genetic element. An oligonucleotide primer bearing the desired mutated sequence, synthetically prepared, is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions. The hybridized product is subjected to DNA polymerizing enzymes such as *E. coli* polymerase I (Klenow fragment) in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed, wherein one strand encodes the original non-mutated sequence, and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate host cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

Comprehensive information on the functional significance and information content of a given residue of protein can best be obtained by saturation mutagenesis in which all 19 amino acid substitutions are examined. The shortcoming of this approach is that the logistics of multiresidue saturation mutagenesis are daunting (Warren *et al.*, 1996, Brown *et al.*, 1996; Zeng *et al.*, 1996; Burton and Barbas, 1994; Yelton *et al.*, 1995; Jackson *et al.*, 1995; Short *et al.*, 1995; Wong *et al.*, 1996; Hilton *et al.*, 1996). Hundreds, and possibly even thousands, of site specific mutants must be studied. However, improved techniques make production and rapid screening of mutants much more straightforward. See also, U.S. Patents 5,798,208 and 5,830,650, for a description of "walk-through" mutagenesis.

Other methods of site-directed mutagenesis are disclosed in U.S. Patents 5,220,007; 5,284,760; 5,354,670; 5,366,878; 5,389,514; 5,635,377; and 5,789,166.

Methods of Making Transgenic Mice

A particular embodiment of the present invention provides transgenic animals that contain the knockin/fusion constructs of interest. In a general aspect, a transgenic animal is produced by the integration of a given transgene into the genome in a manner that permits the expression of the transgene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Patent 4,873,191; which is incorporated herein by

reference), Brinster et al. 1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A Laboratory Manual" 2nd edition (eds., Hogan, Beddington, Costantimi and Long, Cold Spring Harbor Laboratory Press, 1994; which is incorporated herein by reference in its entirety).

Typically, a gene flanked by genomic sequences is transferred by microinjection into a fertilized egg. The microinjected eggs are implanted into a host female, and the progeny are screened for the expression of the transgene. Transgenic animals may be produced from the fertilized eggs from a number of animals including, but not limited to reptiles, amphibians, birds, mammals, and fish.

DNA clones for microinjection can be prepared by any means known in the art. For example, DNA clones for microinjection can be cleaved with enzymes appropriate for removing the bacterial plasmid sequences, and the DNA fragments electrophoresed on 1% agarose gels in TBE buffer, using standard techniques. The DNA bands are visualized by staining with ethidium bromide, and the band containing the expression sequences is excised. The excised band is then placed in dialysis bags containing 0.3 M sodium acetate, pH 7.0. DNA is electroeluted into the dialysis bags, extracted with a 1:1 phenol:chloroform solution and precipitated by two volumes of ethanol. The DNA is redissolved in 1 ml of low salt buffer (0.2 M NaCl, 20 mM Tris,pH 7.4, and 1 mM EDTA) and purified on an Elutip-D™column. The column is first primed with 3 ml of high salt buffer (1 M NaCl, 20 mM Tris, pH 7.4, and 1 mM EDTA) followed by washing with 5 ml of low salt buffer. The DNA solutions are passed through the column three times to bind DNA to the column matrix. After one wash with 3 ml of low salt buffer, the DNA is eluted with 0.4 ml high salt buffer and precipitated by two volumes of ethanol. DNA concentrations are measured by absorption at 260 nm in a UV spectrophotometer. For microinjection, DNA concentrations are adjusted to 3 mg/ml in 5 mM Tris, pH 7.4 and 0.1 mM EDTA.

Other methods for purification of DNA for microinjection are described in Hogan et al. Manipulating the Mouse Embryo (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1986), in Palmiter et al. Nature 300:611 (1982); in The Qiagenologist, Application Protocols, 3rd edition, published by Qiagen, Inc., Chatsworth, CA.; and in Sambrook et al. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).

In an exemplary microinjection procedure, female mice six weeks of age are induced to superovulate with a 5 IU injection (0.1 cc, ip) of pregnant mare serum gonadotropin (PMSG; Sigma) followed 48 hours later by a 5 IU injection (0.1 cc, ip) of human chorionic gonadotropin (hCG; Sigma). Females are placed with males immediately after hCG injection. Twenty-one hours after hCG injection, the mated females are sacrificed by CO2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts and placed in Dulbecco's phosphate buffered saline with 0.5% bovine serum albumin (BSA; Sigma). Surrounding cumulus cells are removed with hyaluronidase (1 mg/ml). Pronuclear embryos are then washed and placed in Earle's balanced salt solution containing 0.5 % BSA (EBSS) in a 37.5°C incubator with a humidified atmosphere at 5% CO2, 95% air until the time of injection. Embryos can be implanted at the two-cell stage.

Randomly cycling adult female mice are paired with vasectomized males. C57BL/6 or Swiss mice or other comparable strains can be used for this purpose. Recipient females are mated at the same time as donor females. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5 % avertin per gram of body weight. The oviducts are exposed by a single midline dorsal incision. An incision is then made through the body wall directly over the oviduct. The ovarian bursa is then torn with watchmakers forceps. Embryos to be transferred are placed in DPBS (Dulbecco's phosphate buffered saline) and in the tip of a transfer pipet (about 10 to 12 embryos). The pipet tip is inserted into the infundibulum and the embryos transferred. After the transfer, the incision is closed by two sutures.

Dosage and Formulation

The compounds (active ingredients) of this invention can be formulated and administered to treat a variety of disease states by any means that produces contact of the active ingredient with the agent's site of action in the body of a mammal. They can be administered by any conventional means available for use in conjunction with pharmaceuticals, either as individual therapeutic active ingredients or in a combination of therapeutic active ingredients. They can be administered alone, but are generally administered with a pharmaceutical carrier selected on the basis of the chosen route of administration and standard pharmaceutical practice.

The dosage administered will be a therapeutically effective amount of active ingredient and will, of course, vary depending upon known factors such as the pharmacodynamic characteristics of the particular active ingredient and its mode and route of administration; age, sex, health and weight of the recipient; nature and extent of symptoms; kind of concurrent treatment, frequency of treatment and the effect desired.

The active ingredient can be administered orally in solid dosage forms such as capsules, tablets and powders, or in liquid dosage forms such as elixirs, syrups, emulsions and suspensions. The active ingredient can also be formulated for administration parenterally by injection, rapid infusion, nasopharyngeal absorption or dermoabsorption. The agent may be administered intramuscularly, intravenously, or as a suppository.

Gelatin capsules contain the active ingredient and powdered carriers such as lactose, sucrose, mannitol, starch, cellulose derivatives, magnesium stearate, stearic acid, and the like. Similar diluents can be used to make compressed tablets. Both tablets and capsules can be manufactured as sustained release products to provide for continuous release of medication over a period of hours. Compressed tablets can be sugar coated or film coated to mask any unpleasant taste and protect the tablet from the atmosphere, or enteric coated for selective disintegration in the gastrointestinal tract.

Liquid dosage forms for oral administration can contain coloring and flavoring to increase patient acceptance.

In general, water, a suitable oil, saline, aqueous dextrose (glucose), and related sugar solutions and glycols such as propylene glycol or polyethylene glycols are suitable carriers for parenteral solutions. Solutions for parenteral administration contain preferably a water soluble salt of the active ingredient, suitable stabilizing agents and, if necessary, buffer substances. Antioxidizing agents such as sodium bisulfate, sodium sulfite or ascorbic acid, either alone or combined, are suitable stabilizing agents. Also used are citric acid and its salts and sodium Ethylenediaminetetraacetic acid (EDTA). In addition, parenteral solutions can contain preservatives such as benzalkonium chloride, methyl- or propyl-paraben and chlorobutanol. Suitable pharmaceutical carriers are described in *Remington's Pharmaceutical Sciences*, a standard reference text in this field.

Additionally, standard pharmaceutical methods can be employed to control the duration of action. These are well known in the art and include control release preparations

and can include appropriate macromolecules, for example polymers, polyesters, polyamino acids, polyvinyl, pyrolidone, ethylenevinylacetate, methyl cellulose, carboxymethyl cellulose or protamine sulfate. The concentration of macromolecules as well as the methods of incorporation can be adjusted in order to control release. Additionally, the agent can be incorporated into particles of polymeric materials such as polyesters, polyamino acids, hydrogels, poly (lactic acid) or ethylenevinylacetate copolymers. In addition to being incorporated, these agents can also be used to trap the compound in microcapsules.

Useful pharmaceutical dosage forms for administration of the compounds of this invention can be illustrated as follows.

Capsules: Capsules are prepared by filling standard two-piece hard gelatin capsulates each with 100 milligram of powdered active ingredient, 175 milligrams of lactose, 24 milligrams of talc and 6 milligrams magnesium stearate.

Soft Gelatin Capsules: A mixture of active ingredient in soybean oil is prepared and injected by means of a positive displacement pump into gelatin to form soft gelatin capsules containing 100 milligrams of the active ingredient. The capsules are then washed and dried.

Tablets: Tablets are prepared by conventional procedures so that the dosage unit is 100 milligrams of active ingredient. 0.2 milligrams of colloidal silicon dioxide, 5 milligrams of magnesium stearate, 275 milligrams of microcrystalline cellulose, 11 milligrams of cornstarch and 98.8 milligrams of lactose. Appropriate coatings may be applied to increase palatability or to delay absorption.

Injectable: A parenteral composition suitable for administration by injection is prepared by stirring 1.5% by weight of active ingredients in 10% by volume propylene glycol and water. The solution is made isotonic with sodium chloride and sterilized.

Suspension: An aqueous suspension is prepared for oral administration so that each 5 millimeters contain 100 milligrams of finely divided active ingredient, 200 milligrams of sodium carboxymethyl cellulose, 5 milligrams of sodium benzoate, 1.0 grams of sorbitol solution U.S.P. and 0.025 millimeters of vanillin.

Accordingly, the pharmaceutical composition of the present invention may be delivered via various routes and to various sites in an animal body to achieve a particular effect (see, e.g., Rosenfeld et al. (1991), supra; Rosenfeld et al., Clin. Res., 39(2), 311A (1991a); Jaffe et al., supra; Berkner, supra). One skilled in the art will recognize that

although more than one route can be used for administration, a particular route can provide a more immediate and more effective reaction than another route. Local or systemic delivery can be accomplished by administration comprising application or instillation of the formulation into body cavities, inhalation or insufflation of an aerosol, or by parenteral introduction, comprising intramuscular, intravenous, peritoneal, subcutaneous, intradermal, as well as topical administration.

The composition of the present invention can be provided in unit dosage form wherein each dosage unit, e.g., a teaspoonful, tablet, solution, or suppository, contains a predetermined amount of the composition, alone or in appropriate combination with other active agents. The term "unit dosage form" as used herein refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of the compositions of the present invention, alone or in combination with other active agents, calculated in an amount sufficient to produce the desired effect, in association with a pharmaceutically acceptable diluent, carrier, or vehicle, where appropriate. The specifications for the unit dosage forms of the present invention depend on the particular effect to be achieved and the particular pharmacodynamics associated with the pharmaceutical composition in the particular host.

These methods described herein are by no means all-inclusive, and further methods to suit the specific application will be apparent to the ordinary skilled artisan. Moreover, the effective amount of the compositions can be further approximated through analogy to compounds known to exert the desired effect.

The following examples are offered by way of example, and are not intended to limit the scope of the invention in any manner.

EXAMPLE 1

Placing the agouti cDNA Under the Control of an Imprinted Promoter

The *Snrpn* gene (SEQ ID NO:12) contains 10 exons within about 22 kb of genomic DNA (Shemer et al., 1997), and encodes two open reading frames, both expressed

exclusively from the paternal chromosome. A small open reading frame potentially encodes a 71 amino acid protein of unknown function and extends from exon 1 to within exon 3. A large open reading frame encodes the 240 amino acid SmN protein and is initiated in exon 4. The agouti cDNA (SEQ ID NO:13) was introduced into the *Snrpn* locus in mice by homologous recombination in embryonic stem (ES) cells. A replacement type vector with positive/negative selection markers was constructed to insert the agouti cDNA into the *Eco*RI site of exon 3 of the *Snrpn* gene leaving a 10-kb genomic region of potential regulatory significance flanking upstream of the knockin site (Figure1a).

Construction of the the targeting vector was as follows. Overlapping lambda phage clones containing the *Snrpn* gene were isolated from a mouse 129Sv/Ev genomic library. To construct the agouti knockin vector, a 0.58 kb *Cla*I-*Bam*HI fragment containing the agouti cDNA was eluted from plasmid p284. This fragment and a 0.46-kb *Bgl*II-*Hin*dIII fragment containing the rabbit β-globin polyadenylation signal were cloned into the pBluescript vector (Stratagene, La Jolla, CA). A neomycin expression cassette with the RNA polymerase II promoter (Soriano et al., 1991) was then introduced downstream of the polyadenylation signal. The final targeting vector consisted of a 3.4 kb *Eco*RI genomic DNA fragment as the 5′ homology region and a 2.7 kb *Sal*I (one end from polylinker of phage clone) genomic DNA fragment as the 3′ homology region flanking the agouti and neomycin cassettes. The HSV-tk expression cassette with the rat β-actin promoter (McMahon and Bradley, 1990) was subsequently ligated downstream of the 3′ homology region. The targeting vector was linearized at a unique *Sal*I site outside of the 5′ region of homology, leaving the plasmid backbone attached to the HSV-tk cassette.

EXAMPLE 2

Generation of agouti knockin mice

Introduction of the agouti cDNA and the rabbit β-globin polyadenylation signal into the exon 3 of the *Snrpn* gene terminates the *Snrpn*/agouti fusion RNA before the large open reading frame, although the large open reading frame is potentially intact. The *Snrpn*/agouti fusion RNA contains the complete small open reading frame of the *Snrpn* gene followed by

the agouti open reading frame. The targeting vector containing a neomycin expression cassette for positive selection and the HSV-tk gene for negative selection was linearized, and 15 μg were electroporated into AB2.2 ES cells derived from the 129Sv/Ev mouse strain. Geneticin (G418 sulfate, 200 μg/mL) and FIAU (0.2 μM) selection was applied 24 hours after plating and maintained for 7-8 days. Colonies resistant to G418 and FIAU were picked, trypsinized, and seeded onto a feeder layer of mitotically inactivated STO cells in 96-well plates. Correctly targeted G418-resistant clones were identified by Southern blot analysis using 5′ and 3′ flanking probes shown in Figure 1a to assure that homologous recombination had occurred at both ends. The 5′ flanking probe was a 0.35 kb *Xba*I-*Sac*I fragment from intron 1 of the *Snrpn* gene, and the 3′ flanking probe was a 1.2-kb *Hin*dIII fragment from intron 3 of the *Snrpn* gene. Sixteen targeted clones were identified after screening 96 G418-resistant colonies. The targeted ES cells were injected into C57BL/6J blastocysts. Male chimeras were bred with C57BL/6J females, and germline transmission of the *Snrpn*/agouti fusion transgene was obtained (Figure1b). The heterozygous progeny with a 129Sv/Ev and C57BL/6J mixed background (Figure 2a, generation I) were further crossed with C57BL/6J mice to transfer the agouti transgene onto the homozygous nonagouti (*a/a*) background (Fig.2a, generation II) as determined by Southern blot analysis to distinguish the agouti *A* and nonagouti *a* alleles (Vrieling et al., 1998). Only mice of *a/a* genotype were selected for further study.

EXAMPLE 3

The agouti knockin/fusion sequence yields an imprinted coat color effect

The effect of the agouti knockin/fusion sequence on coat color is demonstrated in Figure 3. When the agouti knockin/fusion sequence was inherited from a heterozygous female, the offspring carrying the knockin/fusion sequence were entirely black and were indistinguishable from nonagouti *a/a* littermates (Figure 3b). In contrast, when the agouti knockin/fusion sequence was inherited from a heterozygous male, the offspring carrying the knockin/fusion sequence demonstrated a distinctly lighter tannish abdominal coloration that was clearly different from the completely black, nonagouti *a/a* littermates that did not inherit

a knockin/fusion sequence (Figure 3a). The ventral coat color of mice carrying the imprinted agouti knockin/fusion sequence switches back and forth in both directions through gametogenesis. Breeding of the male homozygous for the knockin/fusion sequence with C57BL/6J females (nonagouti *a/a*) gave 100% of the offspring with tannish ventrum. About half (48.1%) of the offspring have tannish ventrum from the breeding of a heterozygous male carrying the maternally inherited agouti knockin/fusion sequence with C57BL/6J (nonagouti *a/a*) females, indicating switching of the agouti knockin/fusion sequence from a maternal methylated to paternal unmethylated pattern during gametogenesis accompanied reexpression of the knockin/fusion sequence. Conversely, for the reciprocal cross using homozygous or heterozygous females carrying the paternally inherited knockin/fusion sequence bred with C57BL/6J (nonagouti *a/a*) males, 0% of the offspring showed tannish ventrum, indicating that the actively expressed paternal knockin/fusion sequence was switched to a maternal expression pattern and silenced during gametogenesis in the female. In summary, 68 mice inheriting the knockin/fusion sequence paternally had a tannish abdomen, while 69 mice inheriting the knockin/fusion sequence maternally had a black abdomen.

EXAMPLE 4

Switching of methylation pattern is preserved

There is a differentially methylated region at the 5′ end of the mouse *Snrpn* gene extending about 6 kb from promoter to intron 1, including the *Snrpn* CpG island, which correlates inversely with *Snrpn* expression (Shemer et al., 1997). The maternal allele of the *Snrpn* gene is methylated at the CpG island and silenced, while the paternal allele is unmethylated and actively expressed. Figure 2 shows an analysis of the methylation status of the 5′ differentially methylated sites (*Sac*II and *Bss*HII within intron 1 of the *Snrpn* gene; (Figure 1a) in three generations of mice carrying the agouti knockin/fusion sequence. Genomic DNA was isolated by the proteinase K/SDS digestion and phenol/chloroform extraction method (Sambrook et al., 1989). Genomic DNA (10μg) from mice I-1, II-2, and III-2 (Figure 2a) as well as the 129Sv/Ev control was digested overnight with *Hin*dIII alone

or in combination with methylation-sensitive enzymes, *Sac*II or *Bss*HII. Digested DNA was transferred from 0.8% of agarose gels to Hybond N+ membrane (Amersham, Newark, NJ) as described (Sambrook et al., 1989) for Southern blot analysis. The probe was a 1.3-kb *Bss*HII-*EcoR*I fragment from the *Snrpn* intron 1 region. Radiolabelled DNA probe was synthesized using the random hexamer method (Zoon, 1987), and hybridized in a buffer containing 0.25 M $NaPO_4$ pH 7, 0.25 M NaCl, 1 mM EDTA, 10% PEG-8000, 7%SDS, and 1% bovine serum albumin. Probes containing repetitive sequences were preassociated using an excess of mouse DNA to quench repeat hybridization (Sealey et al., 1985). Filters were washed to a final stringency of 0.5X SSC/0.1%SDS at 65 C. Autoradiography was performed at -80 C with intensifying screen and Kodak XA-R film.

As shown in Figure 2b, the 12-kb *Hin*dIII genomic fragment was completely digested by *Sac*II and *Bss*HII on the paternally inherited agouti knockin/fusion sequence in mouse I-1, while the maternally inherited wild type *Snrpn* allele was methylated and resistant to digestion with the methylation-sensitive enzymes (Figure 2b). In contrast, when the paternally inherited agouti knockin/fusion sequence was transmitted through a female, the agouti knockin/fusion sequence became fully methylated (Figure 2b, mouse II-2), indicating that there was a switch of the methylation pattern of *Snrpn* CpG island from a paternal to maternal state. When the maternally inherited agouti knockin/fusion sequence was further transmitted through a male, the methylation of the *Snrpn* CpG island switched back to the paternal unmethylated status (Figure 2b,mouse III-2). Thus, the differential methylation of the *Snrpn* CpG island switches from a paternal to the maternal methylated states or maternal to the paternal unmethylated state as would be expected in normal mice.

EXAMPLE 5

The imprinting of expression is retained for the mutant allele

RT-PCR analysis was used to examine the RNA expression of the wild type *Snrpn* allele and the *Snrpn*/agouti knockin/fusion sequence. Total RNA was isolated using the guanidinium thiocyanate/CsCl gradient method (Sambrook et al., 1989). For RT-PCR analysis, DNaseI-treated total RNA was reverse transcribed using primer p3 (5'-

CTGTTCCACAATAGCCGTTGTC-3'; SEQ ID NO;1) for the *Snrpn* wild type transcripts, p5 (5'-GCGAGTTCATGGAGGAGTTACTCC-3'; SEQ ID NO:2) for the *Snrpn*/agouti knockin/fusion sequence, primer Neo-3 (5'-TGATGCTCTTCGTCCAGATCATCC-3'; SEQ ID NO:3) for neomycin transgene, and primer HPRT-3 (5'-CTTTCCAGTTAAAGTTGAGAGATC-3'; SEQ ID NO:4) for the HPRT gene. All PCR was carried out in 50 µl reactions containing 10 mM Tris pH8.3, 1.5 mM MgCl2, 50 mM KCl, 200 µM dNTPs, 1µM primers, and 2.5 µl of RT reaction with cycling conditions of 94° C for 1 min, 59° C for 1 min, and 72° C for 1 min for 30 cycles. PCR products were analyzed on a 1.2% agarose gel. The PCR primer sequences are as follows: for *Snrpn* wild type allele, p1 (forward) 5'-TTGGTTCTGAGGAGTGATTTGC-3' (SEQ ID NO:5), and p2 (reverse) 5'-CCTTGAATTCCACCACCTTG-3' (SEQ ID NO:6); for *Snrpn*/agouti knockin/fusion sequence, p1 (forward, listed above), and p4 (reverse) 5'-GTGGACGGTGAAGAAGCACAGGAAG-3' (SEQ ID NO:7); for neomycin transgene, Neo-1 (forward) 5'-CTTTTTGTCAAGACCGACCTGTCCG-3' (SEQ ID NO:8), and Neo-2 (reverse) 5'-CTCGATGCGATGTTTCGCTTGGTG-3' (SEQ ID NO:9); for HPRT internal control, HPRT-1 (forward) 5'-ATGACCTAGATTTGTTTTGTATACC-3' (SEQ ID NO:10), and HPRT-2 (reverse) 5'-GTAGCTCTTCAGTCTGATAAAATCTAC-3'(SEQ ID NO:11).

Figure 4 shows that only the paternally inherited transcript (either the *Snrpn* wild type or *Snrpn*/agouti knockin/fusion sequence) is present in mice heterozygous for the knockin/fusion sequence. The *Snrpn*/agouti knockin/fusion sequence RNA was expressed in the skin of mice I-1 and III-1 (see pedigree in Figure 2a) carrying the paternally inherited agouti knockin/fusion sequence. When the knockin/fusion sequence was maternally inherited, the *Snrpn*/agouti knockin/fusion sequence was silenced, and the *Snrpn* RNA was expressed from the paternal wild type allele (Figure 4, mouse II-2). Imprinted expression of the *Snrpn*/agouti knockin/fusion sequence was also detected using RNA from the brain of mouse III-1. RNA containing the large open reading frame of *Snrpn* from the agouti knockin/fusion sequence was not detected by RT-PCR analysis of *Snrpn* exon 7. Therefore, the RT-PCR analysis shows that RNA is expressed from the paternal unmethylated chromosome but not from the maternal methylated chromosome. Taken together, the results demonstrate coat color variation based on imprinted regulation of the agouti cDNA under the control of the *Snrpn* promoter has been achieved.

EXAMPLE 6

Demonstration of the Feasability of a Screen for Identification of Genetic Alterations That Affect Genomic Imprinting of the *Snrpn* Promoter

To further dissect the regulatory mechanisms controlling genomic imprinting, a novel strategy has been developed to use mutagenesis in the mouse to identify mutations affecting genomic imprinting. Towards this goal, a coat color marker has been placed under the control of an imprinted promoter so that mutations affecting imprinting can be detected through coat color variation.

Genomic imprinting on human chromosome 15q11-q13 and the homologous region of mouse chromosome 7c as related to Prader Willi and Angelman syndrome, two mental retardation conditions affecting humans, was studied. The mouse and human genomes for this region show considerable similarity. Mutations affecting genomic imprinting are isolated. The *Snrpn* promoter in the mouse is used in the isolation procedure. This promoter lies within a CpG island with methylation and silencing of the maternal chromosome as opposed to absence of methylation and expression on the paternal chromosome. Agouti cDNA was introduced using homologous recombination in embryonic stem (ES) cells, and placed under the control of the *Snrpn* promoter. This knockin/fusion sequence was transmitted to the mouse germline. Coat-color variation was used as an easily observable marker to identify genetic alterations that affect imprinting of the *Snrpn* promoter. In non-agouti (*a/a*) mice, paternal inheritance of the knockin/fusion sequence conferred a tannish abdominal color; the abdomen was black if the knockin/fusion sequence was absent or was inherited maternally when using appropriate genotypes. The differential methylation and imprinted expression of the *Snrpn* promoter was preserved, and imprinting was also preserved when tested at an RNA level. The feasibility of using a knockin/fusion sequence approach for screening in the mouse was demonstrated using two breeding schemes. The two breeding schemes are depicted in Figures 5a and 5b. In breeding scheme 1, a male mouse homozygous for the knockin/fusion sequence and *a/a* were mutagenized (treated with ethylnitrosourea (ENU)) and bred to *a/a* wild type female mouse. All offspring received the knockin/fusion sequence from the mutagenized male parent and had a tannish abdomen. Mutagenesis of mice was performed by methods well known

in the art, such as described in Justice *et al.*, 1999 and Weber *et al.*, 2000. The types of mutations that might be found are discussed herein.

Results of mutagenesis through May 30, 2000 are indicated in Figure 5, with recovery of one mutant in 1832 mice screened with ENU mutagenesis of the wild type male and 0 out of 377 mice screened with mutagenesis of the homozygous transgenic male. The mutagenized male had been mated with a female homozygous for the fusion transgene, and all offspring have a black abdomen unless a mutation occurs. The recovered mutant with a tan abdomen is depicted in Figure 6.

This mutant mouse lacks methylation of the maternally inherited knockin/fusion gene in the mutant mouse, as shown in Figure 7. Southern blotting was performed using the same probe and restriction enzyme combination as for Figure 2 by standard methods in the art. Tail DNA was isolated and digested with *Hin*dIII (H) or *Hin*dIII+ *Sac*II (H/S) as indicated. Digestion with *Hin*dIII gives a larger fragment for the wild type chromosome and a smaller fragment for the knockin/fusion allele, and combined digestion with *Hin*dIII and methylation-sensitive *Sac*II gives clear indication as to whether a particular allele is methylated or unmethylated. Six mice were analyzed including the parental mice (ENU mutagenized wild type male and homozygous transgenic female) and four offspring (three nonmutant littermates and the mutant animal shown as a solid symbol in that the far right). The data for *Snrpn* demonstrate that the male parent is homozygous wild type and has one methylated and one unmethylated chromosome. The female parent is homozygous mutant with again one methylated and one unmethylated chromosome. The three nonmutant littermates have the genotype as expected, all being heterozygous for the knockin mutation and having the wild type allele unmethylated and the mutant allele methylated. The mutant animal is again heterozygous for the knockin mutation and a normal allele, but in this case both alleles are unmethylated and digested by *Sac*II. This confirms that the mutant animal definitively has a molecular abnormality involving imprinting of the *Snrpn* locus. Studies at additional loci indicate that there was also a partial alteration in the methylation at the necdin (*Ndn)* locus. For all of the animals except the mutant, the H/S digestion gives two bands of equal intensity. In the mutant animal the lower band is much more intense and the upper band less intense, indicating that there is partial demethylation of the necdin locus, which is approximately 1 megabase away on the same chromosome. The H19 gene is located at a great distance on the same chromosome

and is part of a different imprinted gene cluster and shows no abnormality of methylation. Similarly the *IgF2R* locus shows normal imprinting and methylation.

Figure 8 shows preparation of bicistronic agouti and GFP reporter sequences under control of the *Snrpn* promoter. In this instance, the plasmid pIRES-EGFP (Clontech; Palo Alto, CA) was utilized. The agouti (Ag) cDNA was placed under the control of the *Snrpn* promoter, as for the construct shown in Figure 1. This was followed by an internal ribosome entry site (IRES) and a cassette encoding enhanced green fluorescent protein (EGFP). A neo resistance cassette was used to introduce the construct into the germline, and the Cre recombinase was then used to delete the neo cassette, by standard methods in the art. This yields a bicistronic reporter cassette expressing agouti and EGFP. This construct is then introduced into mouse ES cells and is subsequently injected for germline transmission.

Figure 9 illustrates the generation of two fusion gene constructs expressing human α-fetoprotein under control of the promoter for the LDL receptor (abbreviations are as follows: SA, splicing acceptor; pA, SV40 polyA signal). In the upper construct depicted in panel A, also referred to as the 5′ targeting construct, the human α-fetoprotein (hAFP) is expressed using an IRES with the construct introduced in the first intron of the LDL receptor locus. In the lower construct in panel A, also referred to as the 3′ targeting construct, the expression cassette is introduced in the 3′ untranslated region of the LDL receptor locus. In the case of the 5′ -targeting construct, exon 1 is spliced to the IRES, and hAFP is then expressed under the control of the promoter for the LDL receptor. In the case of the 3′ -targeting construct, the gene is spliced in the usual mode and is expressed with production of protein from the LDL receptor. The last exon is modified to include the IRES and hAFP and gives rise to expression from the placement in the 3′ -UTR. With the 5′-targeting construct the gene for the LDL receptor is disrupted, while it retains a functional manner with the 3′-targeting construct.

Panel B of Figure 9 demonstrates detectable hAFP in the tissue culture medium for each of the two constructs. The 5′- and 3′-targeting constructs were electroporated into mouse ES cells, and clones were screened for homologous recombination events by standard methods in the art. Appropriate clones were identified for both the 5′- and 3′-targeting constructs. These clones were then grown in tissue culture medium, and the medium was harvested for quantitation of hAFP using a commercial ELISA. The wild type ES cells (WT) do not express any detectable hAFP as expected. Easily detectable levels of hAFP are found with both the

5'-targeting and 3'-targeting constructs at 27 and 50 hours after subculture. The expression level is higher with the 5'-targeting construct. The ES cells are injected for transmission to the mouse germline.

EXAMPLE 7

Bicistronic and more complex reporter sequences

Bicistronic (two reporter sequences) reporter sequences containing combinations such as agouti coat-color marker and human growth hormone or agouti coat-color marker and green fluorescent protein are prepared. A major advantage of a bicistronic reporter sequence is that events irrelevant to the expression of the knockin/fusion sequence are easily distinguished in that they would ordinarily affect one reporter sequence, but not the other. The simultaneous increase or decrease of expression of both reporter sequences in a bicistronic construct greatly facilitates the elimination of uninteresting events from consideration. These bicistronic constructs are prepared by homologous recombination in ES cells including putting the agouti coat-color marker and GFP, and another construct with the agouti coat-color marker and HGH, under the control of the mouse specific nucleotide sequence promoter, for example the *Snrpn* promoter. There are other circumstances where it might be advantageous to have multiple secreted proteins in a bicistronic marker. In cases such as this, combinations of HGH (SEQ ID NO:14), human-AFP (SEQ ID NO:15), rat-AFP (SEQ ID NO:16) or other secreted proteins are utilized. This is particularly relevant to circumstances where a small number of cells in a remote tissue such as brain is secreting the marker protein into the bloodstream. It is important that bicistronic constructs use a separate initiation for protein translation. This is achieved using an internal ribosome entry site (IRES) so that the two reporter proteins are not covalently attached and one reporter sequence is not secondarily affected posttranslationally by effects on the other reporter sequence. In the bicistronic constructs for analysis of transcriptional regulation, the promoter of a specific nucleotide sequence of interest is linked to the bicistronic reporter construct using homologous recombination in ES cells or in any other cultured mammalian or other

cell system or cell extract. Further redundancy is introduced into the system by preparation of more complex tricistronic reporter sequences.

EXAMPLE 8

Transcriptional regulation.

For examination of transcriptional regulation a promoter of interest is used and the reporter sequence is placed under the control of the promoter of interest. Reporter sequences of various types are placed under the control of promoters other than *Snrpn* to demonstrate the generality of this approach. It is well known that regulatory elements can be located near the promoter or at great distances upstream or downstream, including within large introns. In such instances, constructs are designed that mutate the initiation ATG codon and place the reporter sequence further downstream in the gene so that regulated expression of the reporter sequence is achieved. This is tested with the agouti reporter sequence similar to that seen in the Examples above. One skilled in the art recognizes that a variety of strategies can be used to preserve normal transcriptional control while placing the reporter sequence under the control of the promoter of interest.

EXAMPLE 9

Posttranscriptional regulation

The knockin/fusion sequence strategy is also used to study certain types of posttranscriptional regulation. The reporter sequence is fused to the N-terminal or C-terminal portion of a protein to give a hybrid protein. Under these circumstances, a protein that is subject to degradation by a ubiquitin pathway, for example, can have its half-life determined by that mechanism, and the half-life of the fused reporter protein is similarly affected. The knockin/fusion sequences of this type are used to search for posttranscriptional effects. The constructs are introduced by homologous recombination into ES cells and introduced into the germline. Constructs are also introduced into other cell types by homologous recombination

or introduced as extra chromosomal molecules. In the case of posttranscriptional regulatory effects, other antigenic tags can be fused to proteins to facilitate quantitation. This invention combines the use of these protein fusions to screen for genetic effects on posttranscriptional control or biologic affects of compounds on posttranscriptional control.

EXAMPLE 10

Potential deleterious effects of knockin/fusion sequences on phenotype

The introduction of the agouti cDNA under the control of the *Snrpn* promoter potentially inactivates the *Snrpn* gene. Since it was already known that homozygous knockout of the *Snrpn* gene did not give an obvious mouse phenotype, it was possible to breed animals either heterozygous or homozygous for the knockin/fusion sequence. One skilled in the art recognizes that in many instances, the knockin/fusion sequence is only useful in the heterozygous state because the homozygous inactivation of the specific nucleotide sequence of interest is lethal or causes other phenotypic abnormalities. Further, the skilled artisan recognizes that even a heterozygous knockin/fusion sequence introduced by homologous recombination can have a deleterious or even lethal effect. In these instances, constructs are made in which the reporter sequence is introduced with an IRES restart in the 3′ untranslated portion of the specific nucleotide sequence of interest. Using this strategy, any nucleotide sequence can be analyzed using the proposed methodology. Introduction of the reporter sequence into the 3′ untranslated portion of the specific nucleotide sequence of interest using an IRES restart can be the preferred strategy.

EXAMPLE 11

Biological effects of compounds

As has already been demonstrated using 5-azacytidine, it is possible to screen for effects of compounds on sequence expression in this system. Various compounds with the potential to regulate transcription can be tested with the methods of the present invention. These include,

but are not limited to, antisense nucleic acids, small xenotrophic molecules, and known biological agents occurring naturally in cells. Screening of compounds is carried out in vivo with intact animals such as mice or in vitro in cells, tissues, organs or tissue slices. This approach was used in the administration of azacytidine to intact animals. Tissue culture cells carrying the knockin/fusion sequence on the maternal chromosome with it methylated and silenced are available. With a bicistronic reporter sequence including agouti cDNA and human growth hormone, the secretion of human growth hormone is induced by treatment of the fibroblast cultures with 5-azacytidine. This demonstrates the feasibility of screening compounds for biological effects either in vivo or in vitro in tissue culture. Although large numbers of compounds can be administered to animals such as mice to examine for obvious effects on coat color or increases or decreases in the blood level of HGH, tissue culture allows high-throughput screening. For example, 96 well tissue culture dishes can be used to screen multiple compounds easily and quickly with the methods described herein.

In the instance of drugs, bicistronic tests are available. For example, a bicistronic message expressing HGH and GFP is placed under the control of the utrophin promoter (SEQ ID NO:17). There is ample evidence that increased expression of utrophin in muscle can provide a cure for Duchenne muscular dystrophy, where dystrophin is genetically defective or absent. A large number of compounds are screened in intact mice and/or in cultured cells derived from such mice. The compounds that induce increased expression of utrophin can cause increased levels of growth hormone in the blood or tissue culture medium and thus induce the histological detectability of GFP. This is an example representing a search for upregulators of gene expression.

A bicistronic reporter expressing HGH and hAFP under the control of the amyloid A beta precursor protein (APP) (SEQ ID NO:18) promoter is tested. An alternative embodiment is placement of the bicistronic reporter sequence under the control of the promoter for the presenilin-1 sequence (SEQ ID NO:19) or presenilin-2 sequence (SEQ ID NO:20), which likely interfere with processing of APP, or the α-secretase (SEQ ID NO:21) that cleaves amyloid A in a manner that promotes Alzheimer disease. These genes are expressed at detectable levels in tissue culture cells. Again, intact animals such as mice or cell cultures derived from humans are tested. With the bicistronic HGH and hAFP reporter sequences, compounds that increase the expression of either of these genes show increased levels of HGH and hAFP in blood or

tissue culture medium and drugs that decrease expression show decreased levels of HGH and hAFP in blood or tissue culture medium. These two genes are tested separately seeking decreased expression as a preventive approach in Alzheimer disease.

EXAMPLE 12

Additional Specific Embodiments

The methods described herein may be adapted to other sequences, promoters, cell lines, systems, etc. in accordance with the teachings provided herein. Additional specific embodiments include a mouse cell line for screening drugs which affect the LDL receptor (SEQ ID NO:22). Alternatively, apolipoprotein A-I (SEQ ID NO:23) is utilized as a nucleic acid sequence of interest, wherein a system is created to screen for drugs that upregulate expression of this nucleic acid, which increases the level of HDL, or beneficial cholesterol. The current lack in the art of effective drugs to raise HDL cholesterol makes this system a valuable addition to the field.

In another embodiment a system is created in accordance with the methods described herein to assay for drugs which regulate expression of genes implicated in pathogenesis of Alzheimers or other neurodegenerative diseases. For example, it is desirable to identify drugs which downregulate the expression of the amyloid beta A4 precursor protein (SEQ ID NO:18) on human chromosome 21. This gene is strongly implicated in the pathogenesis of Alzheimer's and it is thought that three copies of the gene in trisomy 21 contribute to early onset of Alzheimer's disease in that particular medical condition. Other genes of interest would include presenilin-1 (SEQ ID NO:19) and presenilin-2 (SEQ ID NO:20).

REFERENCES

All patents and publications mentioned in the specification are indicative of the level of those skilled in the art to which the invention pertains. All patents and publications are

herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

PATENTS

U.S. Patent No. 5,840,873, issued Nov. 24, 1998
U.S. Patent No. 5,843, 640, issued Dec. 1, 1998
U.S. Patent No. 5,843,650, issued Dec. 1. 1998
U.S. Patent No. 5,843,651, issued Dec. 1, 1998
U.S. Patent No. 5,843,663, issued Dec. 1, 1998
U.S. Patent No. 5,846,708, issued Dec. 8, 1998
U.S. Patent No. 5,846,709, issued Dec. 8, 1998
U.S. Patent No. 5,846,717, issued Dec. 8, 1998
U.S. Patent No.5,846,726, issued Dec. 8, 1998
U.S. Patent No. 5,846,729, issued Dec. 8, 1998
U.S. Patent No. 5,846,783, issued Dec. 8, 1998
U.S. Patent No. 5,849,481, issued Dec. 15, 1998
U.S. Patent No. 5,849,483, issued Dec. 15, 1998
U.S. Patent No. 5,849,486, issued Dec. 15, 1998
U.S. Patent No. 5,849,487, issued Dec. 15, 1998
U.S. Patent No. 5,849,497, issued Dec. 15, 1998
U.S. Patent No. 5,849,546, issued Dec. 15, 1998
U.S. Patent No. 5,849,547, issued Dec. 15, 1998
U.S. Patent No. 5,851,770, issued Dec. 22, 1998
U.S. Patent No. 5,851,772, issued Dec. 22, 1988
U.S. Patent No.5,853,990, issued Dec. 29, 1998
U.S. Patent No. 5,853, 993, issued Dec. 29, 1998
U.S. Patent No. 5,853,992, issued Dec. 29, 1998
U.S. Patent No. 5,856,092, issued Jan. 5, 1999
U.S. Patent No. 5,858,652, issued Jan. 12, 1999
U.S. Patent No. 5,861,244, issued Jan. 19, 1999
U.S. Patent No.5,863,732, issued Jan. 26, 1999

U.S. Patent No. 5,863,753, issued Jan. 26, 1999
U.S. Patent No. 5,866,331, issued Feb. 2, 1999
U.S. Patent No. 5,866,336, issued Feb. 2, 1999
U.S. Patent No. 5,866,337, issued Feb. 2, 1999
U.S. Patent No. 5,900,481, issued May 4, 1999
U.S. Patent No. 5,905,024, issued May 18, 1999
U.S. Patent No. 5,910,407, issued June 8, 1999
U.S. Patent No. 5,912,124, issued June 15, 1999
U.S. Patent No. 5,912,145, issued June 15, 1999
U.S. Patent No. 5,912,148, issued June 15, 1999
U.S. Patent No. 5,916,776, issued June 29, 1999
U.S. Patent No. 5,916,779, issued June 29, 1999
U.S. Patent No. 5,919,626, issued July 6, 1999
U.S. Patent No. 5,919,630, issued July 6, 1999
U.S. Patent No. 5,922, 574, issued July 13, 1999
U.S. Patent No. 5,925,517, issued July 20, 1999
U.S. Patent No. 5,925,525, issued Jul. 20, 1999
U.S. Patent No. 5,928,862, issued July 27, 1999
U.S. Patent No. 5,928,869, issued July 27, 1999
U.S. Patent No. 5,928,870, issued, Jul. 27, 1999
U.S. Patent No. 5,928,905, issued July 27, 1999
U.S. Patent No. 5,928,906, issued July 27, 1999
U.S. Patent No. 5,929,227, issued July 27, 1999
U.S. Patent No. 5,932,413, issued Aug. 3, 1999
U.S. Patent No. 5,932,451, issued Aug. 3, 1999
U.S. Patent No. 5,935,791, issued Aug. 10, 1999
U.S. Patent No. 5,935,825, issued Aug. 10, 1999
U.S. Patent No. 5,939,291, issued Aug. 17, 1999
U.S. Patent No. 5,942,391, issued Aug. 24, 1999
European Application No. 320 308
European Application No. 329 822

GB Application No. 2 202 328

PCT Application No. PCT/US87/00880

PCT Application No. PCT/US89/01025

PCT Application WO 88/10315

PCT Application WO 89/06700

PCT Application WO 90/07641

PUBLICATIONS

Barr, J.A., Jones, J., Glenister, P.H., and Cattanach, B.M. 1995. Ubiquitous expression and imprinting of *Snrpn* in the mouse. *Mamm. Genome* 6: 405-407.

Barsh, G.S. 1996. The genetics of pigmentation: from fancy genes to complex traits. *Trends Genet.* 12: 299-305.

Bartolomei, M.S. and Tilghman, S.M. 1997. Genomic imprinting in mammals. *Annu. Rev. Genet.* 31: 493-525.

Battey, J., Jordan, E., Cox, D., and Dove, W. 1999. An action plan for mouse genomics. *Nat. Genet.* 21: 73-75.

Bultman, S.J., Klebig, M.L., Michaud, E.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a^t) and white-bellied agouti (A^w) reveals alternative forms of agouti transcripts. *Genes Dev.* 8: 481-490.

Bultman, S.J., Michaud, E.J., and Woychik, R.P. 1992. Molecular characterization of the mouse agouti locus. *Cell* 71: 1195-1204.

Cattanach, B.M., Barr, J.A., Evans, E.P., Burtenshaw, M., Beechey, C.V., Leff, S.E., Brannan, C.I., Copeland, N.G., Jenkins, N.A., and Jones, J. 1992. A candidate mouse model for Prader-Willi syndrome which shows an absence of Snrpn expression. *Nat. Genet.* 2: 270-274.

Conroy, J.M., Grebe, T.A., Becker, L.A., Tsuchiya, K., Nicholls, R.D., Buiting, K., Horsthemke, B., Cassidy, S.B., and Schwartz, S. 1997. Balanced translocation 46,XY,t(2;15)(q37.2;q11.2) associated with atypical Prader-Willi syndrome. *Am. J. Hum. Genet.* 61: 388-394.

Dittrich, B., Buiting, K., Korn, B., Rickard, S., Buxton, J., Saitoh, S., Nicholls, R.D., Poustka, A., Winterpacht, A., Zabel, B., and Horsthemke, B. 1996. Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. *Nat. Genet.* 14: 163-170.

Duhl, D.M., Stevens, M.E., Vrieling, H., Saxon, P.J., Miller, M.W., Epstein, C.J., and Barsh, G.S. 1994b. Pleiotropic effects of the mouse lethal yellow (Ay) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. *Development* 120: 1695-1708.

Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L., and Barsh, G.S. 1994a. Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* 8: 59-65.

Fang, P., Lev-Lehman, E., Tsai, T.-F., Matsuura, T., Benton, C.S., Sutcliffe, J.S., Christian, S.L., Kubota, T., Halley, D.J., Meijers-Heijboer, H., Langlois, S., Graham, J.M., Beuten, J., Willems, P., Ledbetter, D.H., and Beaudet, A.L. 1999. The spectrum of mutations in *UBE3A* causing Angelman syndrome. *Hum. Mol. Genet.* 8: 129-135.

Green, M.C. 1989. Genetic Variants and Strains of the Laboratory Mouse. (New York: Oxford University Press).

Hollander, W.F. and Gowen, J.W. 1956. An extreme non-agouti mutant in the mouse. *J. Hered.* 47: 221-224.

Huntriss, J.D., Barr, J.A., Horn, D.A., Williams, D.G., and Latchman, D.S. 1994. Mice lacking *Snrpn* expression show normal regulation of neuronal alternative splicing events. *Mol. Biol. Rep.* 20: 19-25.

Hustad, C.M., Perry, W.L., Siracusa, L.D., Rasberry, C., Cobb, L., Cattanach, B.M., Kovatch, R., Copeland, N.G., and Jenkins, N.A. 1995. Molecular genetic characterization of six recessive viable alleles of the mouse *agouti* locus. *Genetics* 140: 255-265.

Jackson, I.J. 1994. Molecular and developmental genetics of mouse coat color. *Annu. Rev. Genet.* 28: 189-217.

Jiang, Y., Tsai, T.-F., Bressler, J., and Beaudet, A.L. 1998. Imprinting in Angelman and Prader-Willi syndromes. *Curr. Opin. Genet. Dev.* 8: 334-342.

Justice, M.J., Zheng, B., Woychik, R.P., and Bradley, A. 1997. Using targeted large deletions and high-efficiency N-ethyl-N- nitrosourea mutagenesis for functional analyses of the mammalian genome. *Methods* 13: 423-436.

Klebig, M.L., Wilkinson, J.E., Geisler, J.G., and Woychik, R.P. 1995. Ectopic expression of the agouti gene in transgenic mice causes obesity, features of type II diabetes, and yellow fur. *Proc. Natl. Acad. Sci. U. S. A.* 92: 4728-4732.

Kucera, G.T., Bortner, D.M., and Rosenberg, M.P. 1996. Overexpression of an Agouti cDNA in the skin of transgenic mice recapitulates dominant coat color phenotypes of spontaneous mutants. *Dev. Biol.* 173: 162-173.

Kuslich, C.D., Kobori, J.A., Mohapatra, G., Gregorio-King, C., and Donlon, T.A. 1999. Prader-Willi syndrome is caused by disruption of the *SNRPN* gene. *Am. J. Hum. Genet.* 64: 70-76.

Leff, S.E., Brannan, C.I., Reed, M.L., Özçelik, T., Francke, U., Copeland, N.G., and Jenkins, N.A. 1992. Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nat. Genet.* 2: 259-264.

Li, E., Beard, C., and Jaenisch, R. 1993. Role for DNA methylation in genomic imprinting. *Nature* 366: 362-365.

Li, E., Bestor, T.H., and Jaenisch, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69: 915-926.

Lu, D., Willard, D., Patel, I.R., Kadwell, S., Overton, L., Kost, T., Luther, M., Chen, W., Woychik, R.P., Wilkison, W.O., and Cone, R.D. 1994. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. *Nature* 371: 799-802.

Lyko, F., Brenton, J.D., Surani, M.A., and Paro, R. 1997. An imprinting element from the mouse H19 locus functions as a silencer in *Drosophila*. *Nat. Genet.* 16: 171-173.

Lyko, F., Buiting, K., Horsthemke, B., and Paro, R. 1998. Identification of a silencing element in the human 15q11-q13 imprinting center by using transgenic *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 95: 1698-1702.

Malzac, P., Webber, H., Moncla, A., Graham, J.M., Kukolich, M., Williams, C., Pagon, R.A., Ramsdell, L.A., Kishino, T., and Wagstaff, J. 1998. Mutation analysis of *UBE3A* in Angelman syndrome patients. *Am. J. Hum. Genet.* 62: 1353-1360.

McMahon, A.P. and Bradley, A. 1990. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* 62: 1073-1085.

Michaud, E.J., van Vugt, M.J., Bultman, S.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. 1994. Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* 8: 1463-1472.

Miller, M.W., Duhl, D.M., Vrieling, H., Cordes, S.P., Ollmann, M.M., Winkes, B.M., and Barsh, G.S. 1993. Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. *Genes Dev.* 7: 454-467.

Nicholls, R.D., Saitoh, S., and Horsthemke, B. 1998. Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.* 14: 194-200.

Ozçelik, T., Leff, S., Robinson, W., Donlon, T., Lalande, M., Sanjines, E., Schinzel, A., and Francke, U. 1992. Small nuclear ribonucleoprotein polypeptide N (*SNRPN*), an expressed gene in the Prader-Willi syndrome critical region. *Nat. Genet.* 2: 265-269.

Perry, W.L., Hustad, C.M., Swing, D.A., Jenkins, N.A., and Copeland, N.G. 1995. A transgenic mouse assay for agouti protein activity. *Genetics* 140: 267-274.

Ramirez-Solis, R., Davis, A.C., and Bradley, A. 1993. Gene targeting in embryonic stem cells. *Methods Enzymol.* 225: 855

Ramirez-Solis, R., Rivera-Perez, J., Wallace, J.D., Wims, M., Zheng, H., and Bradley, A. 1992. Genomic DNA microextraction: A method to screen numerous samples. *Anal. Biochem.* 201: 331-335.

Reed, M.L. and Leff, S.E. 1994. Maternal imprinting of human *SNRPN*, a gene deleted in Prader-Willi syndrome. *Nat. Genet.* 6: 163-167.

Reik, W. and Walter, J. 1998. Imprinting mechanisms in mammals. *Curr. Opin. Genet. Dev.* 8: 154-164.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning. A Laboratory Manual (Cold Spring Harbor New York: Cold Spring Harbor Laboratory).

Schimenti, J. and Bucan, M. 1998. Functional genomics in the mouse: phenotype-based mutagenesis screens. *Genome Res.* 8: 698-710.

Schulze, A., Hansen, C., Skakkebaek, N.E., Brondum-Nielsen, K., Ledbetter, D.H., and Tommerup, N. 1996. Exclusion of SNRPN as a major determinant of Prader-Willi syndrome by a translocation breakpoint. *Nat. Genet.* 12: 452-454.

Sealey, P.G., Whittaker, P.A., and Southern, E.M. 1985. Removal of repeated sequences from hybridisation probes. *Nucleic Acids Res.* 13: 1905-1922.

Shemer, R., Birger, Y., Riggs, A.D., and Razin, A. 1997. Structure of the imprinted mouse *Snrpn* gene and establishment of its parental-specific methylation pattern. *Proc. Natl. Acad. Sci. U. S. A.* 94: 10267-10272.

Silvers, W.K. 1979. The Coat Colors of Mice (Springer-Verlag).

Siracusa, L.D. 1994. The agouti gene: turned on to yellow. *Trends Genet.* 10: 423-428.

Soriano, P., Montgomery, C., Geske, R., and Bradley, A. 1991. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell* 64: 693-702.

Slauch, J.M. and Silhavy, T.J. 1991. Genetic Fusions as Experimental Tools. *Meth. Enzym.* 204: 213-248.

Sun, Y., Nicholls, R.D., Butler, M.G., Saitoh, S., Hainline, B.E., and Palmer, C.G. 1996. Breakage in the *SNRPN* locus in a balanced 46,XY,t(15;19) Prader-Willi syndrome patient. *Hum. Mol. Genet.* 5: 517-524.

Surani, M.A. 1998. Imprinting and the initiation of gene silencing in the germ line. *Cell* 93: 309-312.

Vrieling, H., Duhl D.M.J., Millar, S.E., Miller, K.A., and Barsh, G.S. 1998. Differences in dorsal and ventral pigmentation result from regional expression fo the mouse agouti gene. *Proc. Natl. Acad. Sci. U. S. A.* 91: 5667-5671.

Warnecke, P.M., Biniszkiewicz, D., Jaenisch, R., Frommer, M., and Clark, S.J. 1998. Sequence-specific methylation of the mouse H19 gene in embryonic cells deficient in the Dnmt-1 gene. *Dev. Genet.* 22: 111-121.

Wolff, G.L., Roberts, D.W., Morrissey, R.L., Greenman, D.L., Allen, R.R., Campbell, W.L., Bergman, H., Nesnow, S., and Frith, C.H. 1987. Tumorigenic responses to lindane in mice: potentiation by a dominant mutation. *Carcinogenesis* 8: 1889-1897.

Woychik, R.P. and Alagramam, K. 1998. Insertional mutagenesis in transgenic mice generated by the pronuclear microinjection procedure. *Int. J. Dev. Biol.* 42: 1009-1017.

Yang, T., Adamson, T.E., Resnick, J.L., Leff, S., Wevrick, R., Francke, U., Jenkins, N.A., Copeland, N.G., and Brannan, C.I. 1998. A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nat. Genet.* 19: 25-31.

Zoon, R.A. 1987. Safety with ^{32}P- and ^{35}S-labeled compounds. *Methods Enzymol.* 152: 25-29.

One skilled in the art readily appreciates that the present invention is well adapted to carry out the objectives and obtain the ends and advantages mentioned as well as those inherent therein. Systems, pharmaceutical compositions, treatments, methods, procedures and techniques described herein are presently representative of the preferred embodiments and are intended to be exemplary and are not intended as limitations of the scope. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention or defined by the scope of the pending claims.

What we claim:

1. A system to determine elements that regulate expression of a nucleotide sequence comprising a reporter sequence inserted within said nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for said nucleotide sequence.

2. The system of Claim 1 wherein the nucleotide sequence is selected from the group of organisms consisting of mammals, non-mammalian animals, plants, insects, aquatic organisms and avian species.

3. The system of Claim 1 wherein two different reporter sequences are inserted within the nucleotide sequence and both controlled by said promoter.

4. The system of Claim 1 wherein the reporter sequence is selected from the group consisting of the agouti gene, αMSH, α-fetoprotein, β-galactosidase, β-glucuronidase, chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), human growth hormone, luciferase, and epitope tag.

5. As a composition of matter a mammal having a functional system of Claim 1 inserted into its genome.

6. As a composition of matter a non-mammalian animal having a functional system of Claim 1 inserted into its genome.

7. As a composition of matter a plant having a functional system of Claim 1 inserted into its genome.

8. As a composition of matter an insect having a functional system of Claim 1 inserted into its genome.

9. As a composition of matter an aquatic organism having a functional system of Claim 1 inserted into its genome.

10. As a composition of matter an avian species having a functional system of Claim 1 inserted into its genome.

11. As a composition of matter a cell line or cell free extract having a functional system of Claim 1 inserted into the cell.

12. A method for identifying an upregulator or downregulator of expression for a specific nucleotide sequence, comprising the steps of:

inserting a test system into a test organism, wherein the test system comprises a reporter sequence inserted into the specific nucleotide sequence under conditions

wherein the expression of the reporter sequence is controlled by the promoter for the specific nucleotide sequence;

delivering a compound to the test organism containing the test system; and

measuring the effect of said compound on said reporter sequence, wherein when said compound increases the expression of said reporter sequence, the compound is an upregulator and when said compound decreases the expression of said reporter sequence, the compound is a downregulator.

13. The method of Claim 12 wherein the reporter sequence is selected from the group consisting of the agouti gene, αMSH, α-fetoprotein, β-galactosidase, β-glucuronidase, chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), human growth hormone, luciferase, and epitope tag.

14. The method of Claim 12 wherein the test organism is selected from the group consisting of a mammal, a non-mammalian animal, a plant, an insect, an aquatic organism and an avian species.

15. The method of Claim 12, wherein the test organism is a cell line, cell free extract, tissue, tissue slice or organ.

16. A method for identifying an upregulator or downregulator of expression of a specific nucleotide sequence comprising the steps of:

inserting a test system into a test organism, wherein the test system comprises a reporter sequence inserted within the specific nucleotide sequence under conditions wherein the expression of the reporter sequence is controlled by the promoter for the specific nucleotide sequence;

mutagenizing the test organism containing the test system; and

measuring the effect of said mutagenesis on expression of said reporter sequence, wherein when the mutant resulting from said mutagenizing increases the expression of said reporter sequence, the mutant is an upregulator and when the mutant resulting from said mutagenizing decreases the expression of said reporter sequence it is a downregulator.

17. The method of Claim 16 further comprising the steps of:

delivering a compound to the mutagenized test organism; and

measuring the effect of said compound on said reporter sequence, wherein when said compound increases the expression of said reporter sequence it is an upregulator and when said compound decreases the expression of said reporter sequence it is a downregulator.

18. The method of Claim 12 or 16 wherein two different reporter sequences are inserted within the nucleotide sequence and both are under the control of said promoter.

19. The method of Claim 16 wherein the test organism is selected from the group consisting of a non-human mammal, a non-mammalian animal, a plant, an insect, an aquatic organism and an avian species.

20. The method of Claim 16, wherein the test organism is a cell line, cell free extract, tissue, tissue slice or organ.

21. A method of identifying an upregulator to treat a disease in which there is a deficiency in expression of a nucleotide sequence comprising the method of Claim 12 or 16.

22. The method of Claim 21 wherein the disease in Duchenne Muscular Dystrophy and the nucleotide sequence is the sequence coding for utrophin.

23. A method of identifying a downregulator to treat a disease in which there is an excess in expression of a nucleotide sequence comprising the method of Claim 12 or 16.

24. The method of claim of 23 where disease is Alzheimer's and the nucleotide sequence is the sequence coding for amyloid A beta precursor protein or presinilin-1.

25. A method of treating a disease in an organism in which the disease is the result of deficiency in expression of a nucleotide sequence comprising the step of introducing a pharmacologically effective dose of an upregulator identified in the method of Claim 12 or 16.

26. A method of treating a disease in an organism in which the disease is the result of an excess in expression of a nucleotide sequence comprising the step of introducing a

pharmacologically effective dose of a downregulator identified in the method of Claim 12 or 16.

27. A method to identify an upregulator or downregulator to enhance the food content of an animal, aquatic organism, avian species or plant comprising the method of Claim 12 or 16.

28. The system of Claim 1, wherein the reporter sequence is introduced into the 3′ untranslated portion of the nucleotide sequence.

29. The method of Claim 12 for identifying compounds that effect posttranscriptional control at the RNA level, wherein the reporter is inserted by homologous recombination or extrachromosomal nucleic acid, and the resulting construct expresses a fusion transcript.

30. The method of Claim 16 for identifying mutant that effects posttranscriptional control at the RNA level wherein the reporter is inserted by homologous recombination or extrachromosomal nucleic acid, and the resulting construct expresses a fusion transcript.

31. The method of Claim 12 for identifying compounds that effect posttranscriptional control at the protein level by inserting the reporter sequence into the nucleotide sequence by homologous recombination or extrachromosomal nucleic acid, wherein such construct expresses a fusion protein.

32. The method of Claim 16 for identifying mutant that effect posttranscriptional control at the protein level by inserting the reporter sequence into the nucleotide sequence by homologous recombination or extrachromosomal nucleic acid, wherein such construct expresses a fusion protein.

33. The system of Claim 1, wherein the reporter sequence is introduced into the nucleotide sequence at a location that expresses the N-terminal portion of a protein.

34. The system of Claim 1, wherein the reporter sequence is introduced into the nucleotide sequence at a location that expresses the C-terminal portion of a protein.

35. A method of treating a disease or condition in an organism by introducing into the organism a pharmacologically effective dose of an upregulator identified in the method of Claim 12 or 16.

36. The method of Claim 35, wherein the upregulator increases the expression of the nucleotide sequence coding for the LDL receptor.

37. The method of Claim 35, wherein the upregulator increases the expression of the nucleotide sequence coding for apolipoprotein A-I.

38. The method of Claim 35, wherein the upregulator increases the expression of the nucleotide sequence coding for a tumor suppressor gene.

39. The method of Claim 1, wherein the upregulator increases the expression of the nucleotide sequence coding for growth hormone.

40. A method of treating a disease or condition in an organism by introducing into the organism a pharmacologically effective dose of a downregulator identified in the method of Claim 12 or 16.

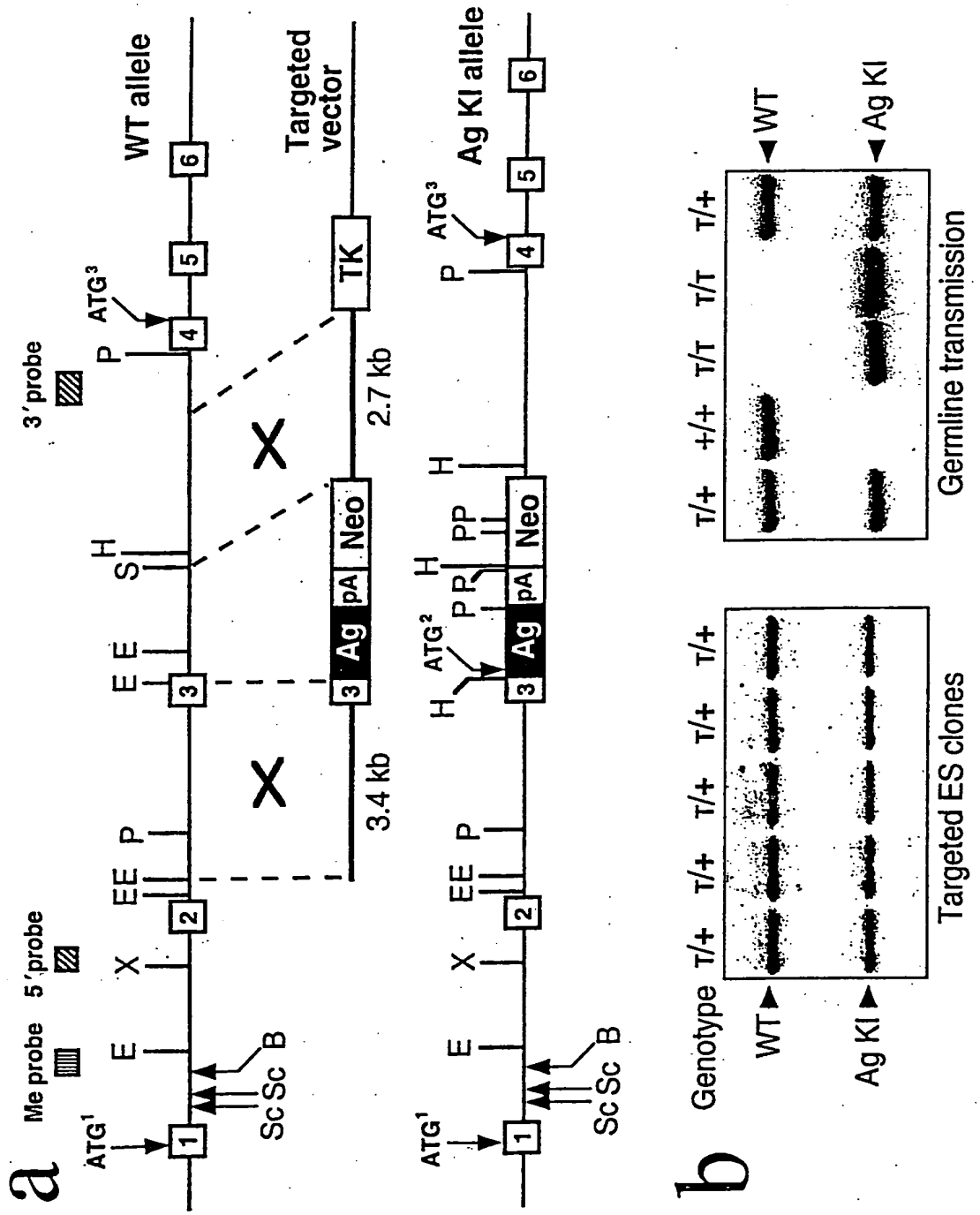

a
WT allele
Targeted vector
Ag KI allele
3' probe
5' probe
Me probe
ATG1
ATG2
ATG3
TK
Neo
pA
Ag
2.7 kb
3.4 kb
E
X
P
S
H
B
Sc
b
Genotype
T/+
+/+
T/T
WT
Ag KI
Targeted ES clones
Germline transmission

Figure 1

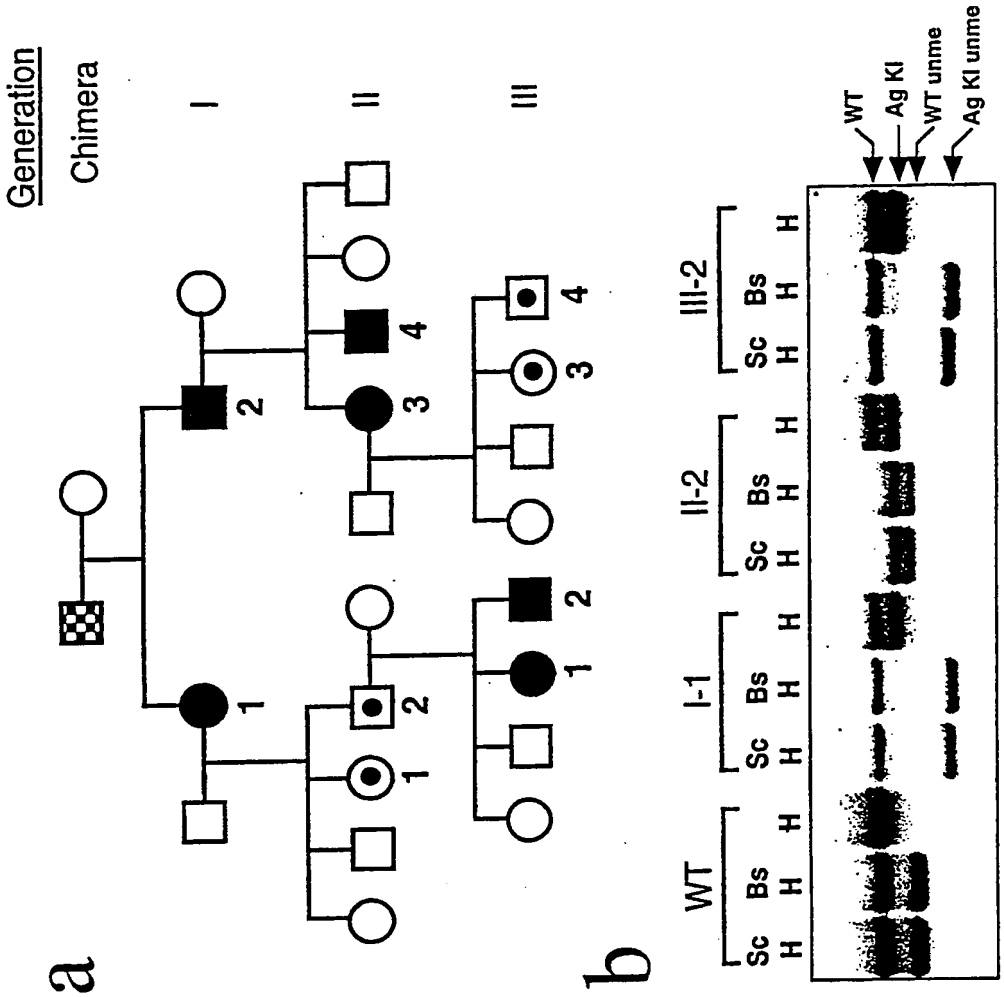

Generation
Chimera
I
II
III
a
b
WT
I-1
II-2
III-2
Sc
Bs
H
WT
Ag KI
WT unme
Ag KI unme

Figure 2

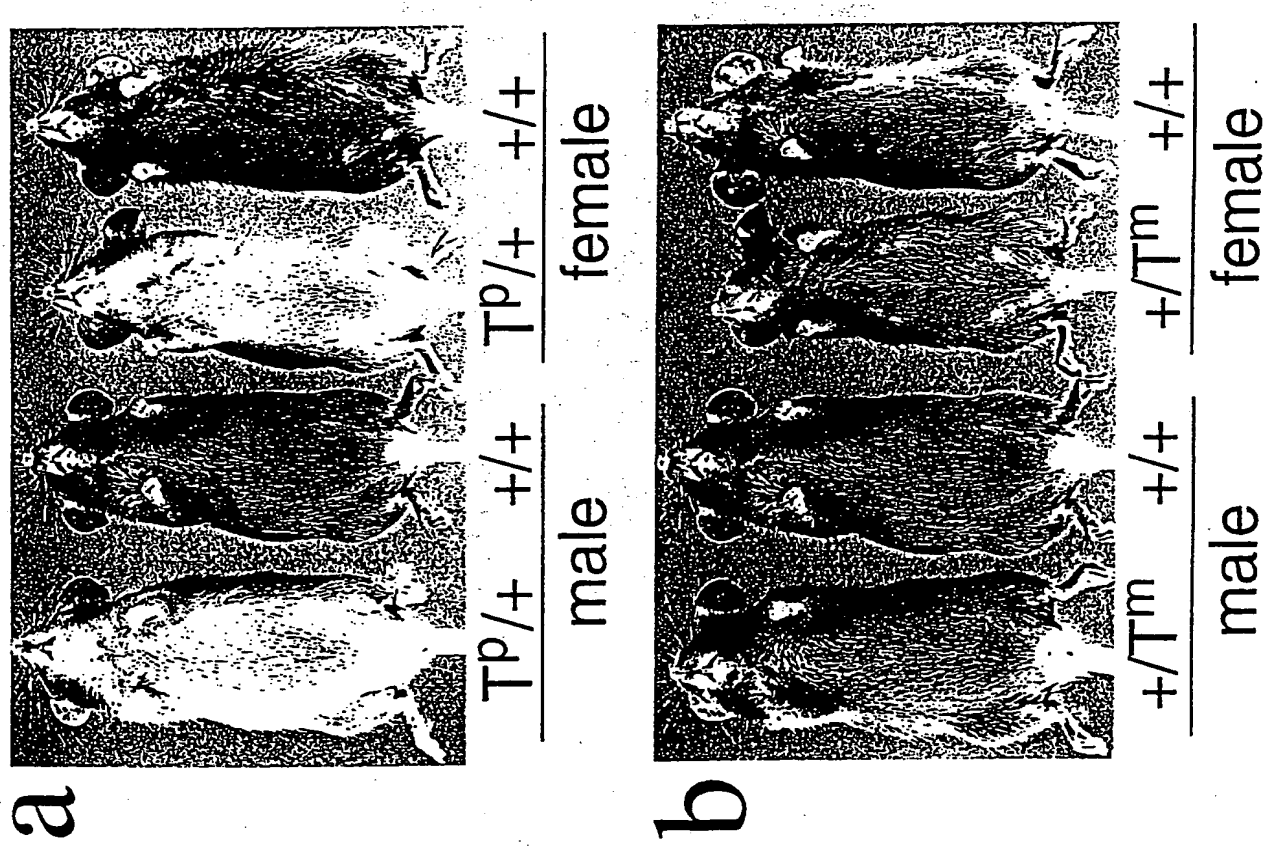


Figure 3

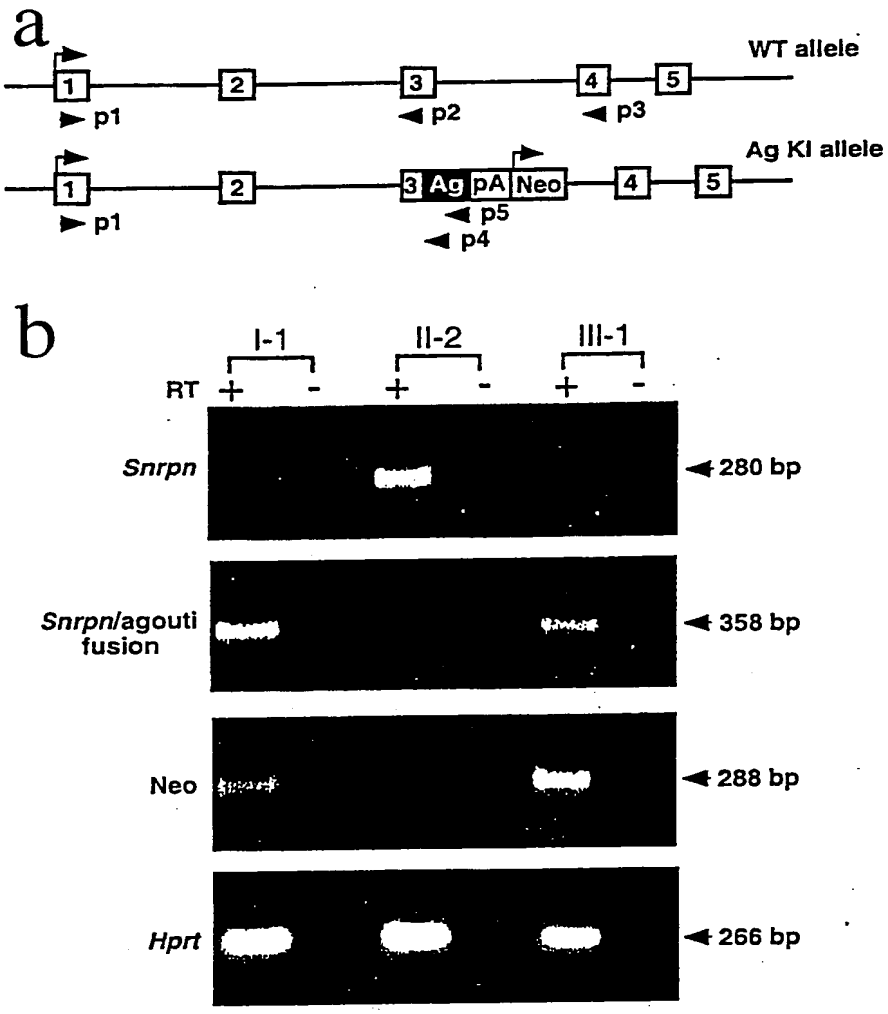


Figure 4

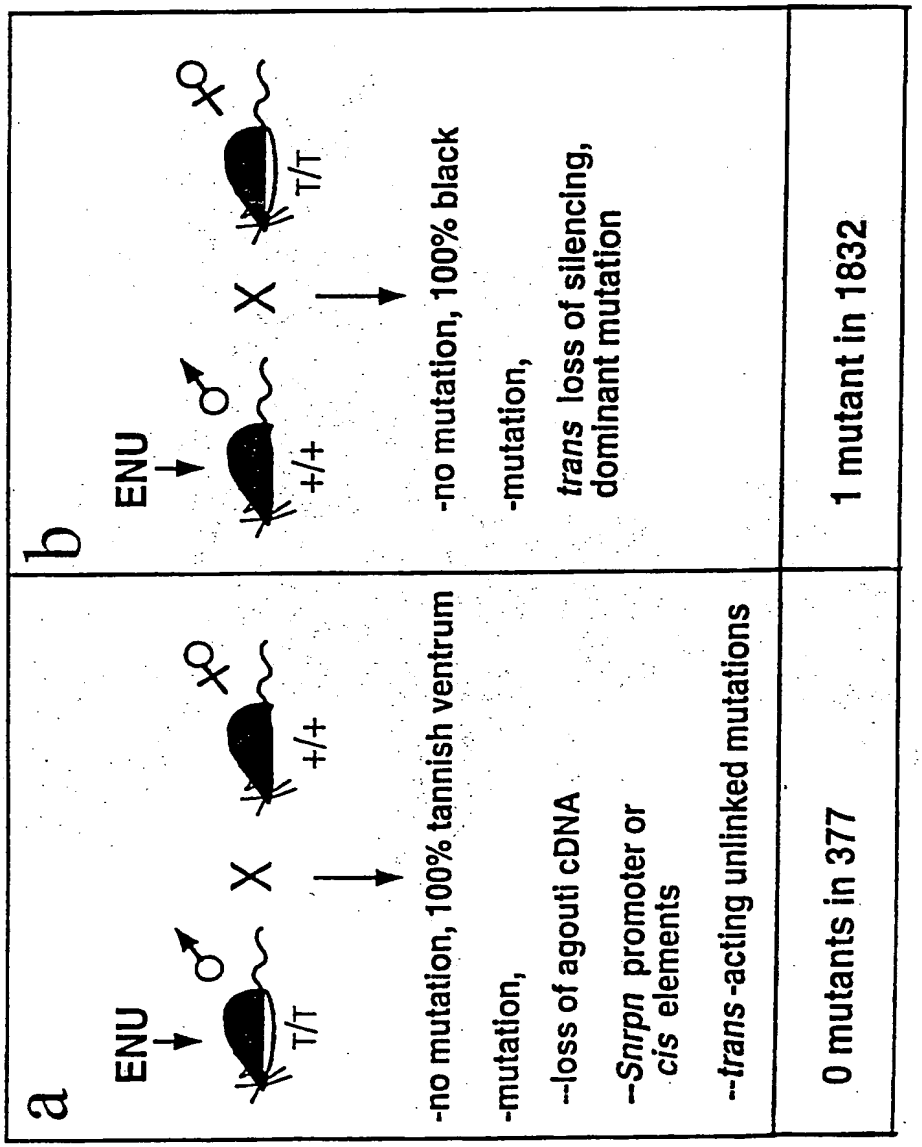

a
ENU
T/T
+/+
-no mutation, 100% tannish ventrum
-mutation,
--loss of agouti cDNA
--Snrpn promoter or cis elements
--trans-acting unlinked mutations
0 mutants in 377
b
ENU
+/+
T/T
-no mutation, 100% black
-mutation,
trans loss of silencing, dominant mutation
1 mutant in 1832

Figure 5

Figure 6

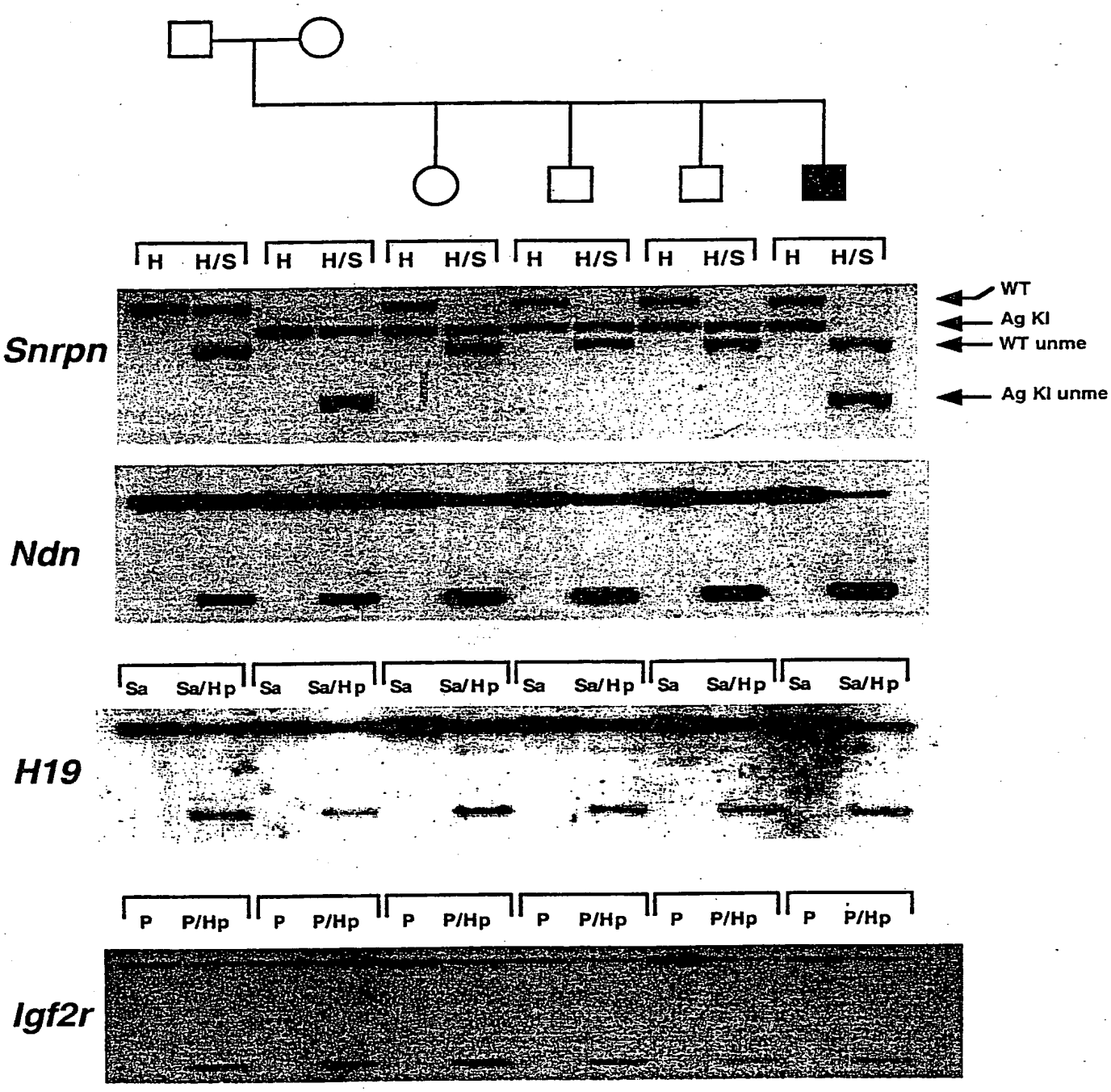


Figure 7

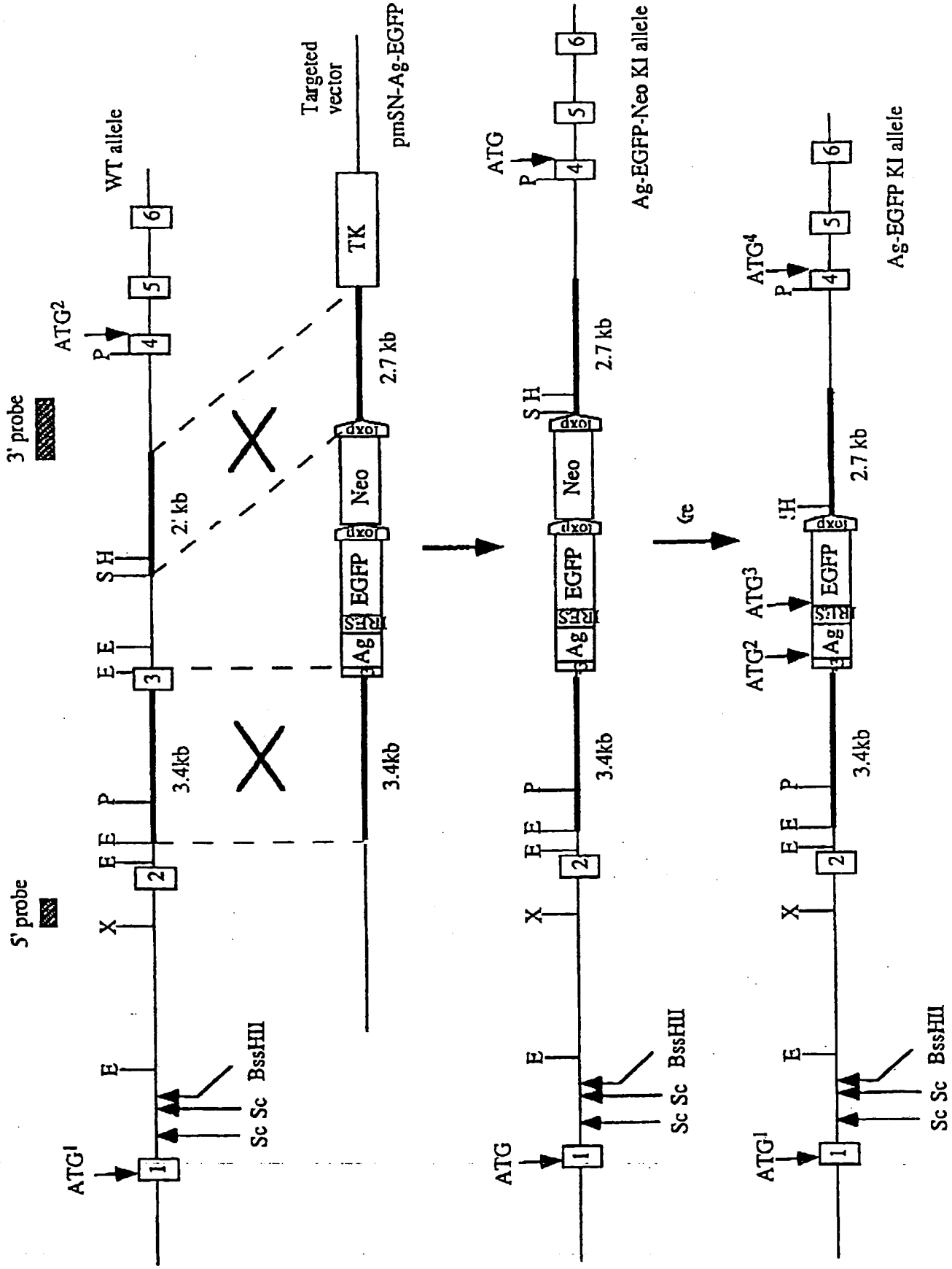
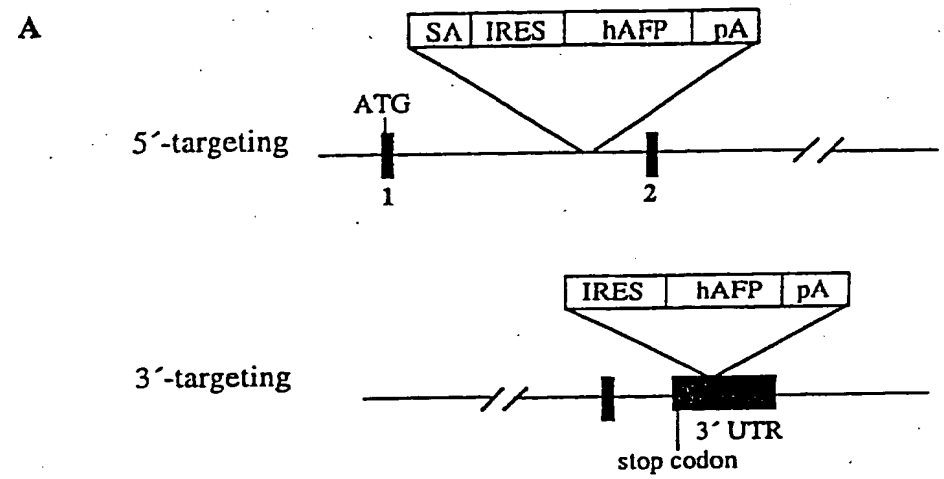

WT allele
Targeted vector
pmSN-Ag-EGFP
Ag-EGFP-Neo KI allele
Ag-EGFP KI allele
3' probe
5' probe
ATG1
ATG2
ATG3
ATG4
TK
Neo
EGFP
Ag
loxP
2.7 kb
3.4kb
2. kb
Sc Sc BssHII

Figure 8


A
SA
IRES
hAFP
pA
ATG
5′-targeting
1
2
IRES
hAFP
pA
3′-targeting
3′ UTR
stop codon

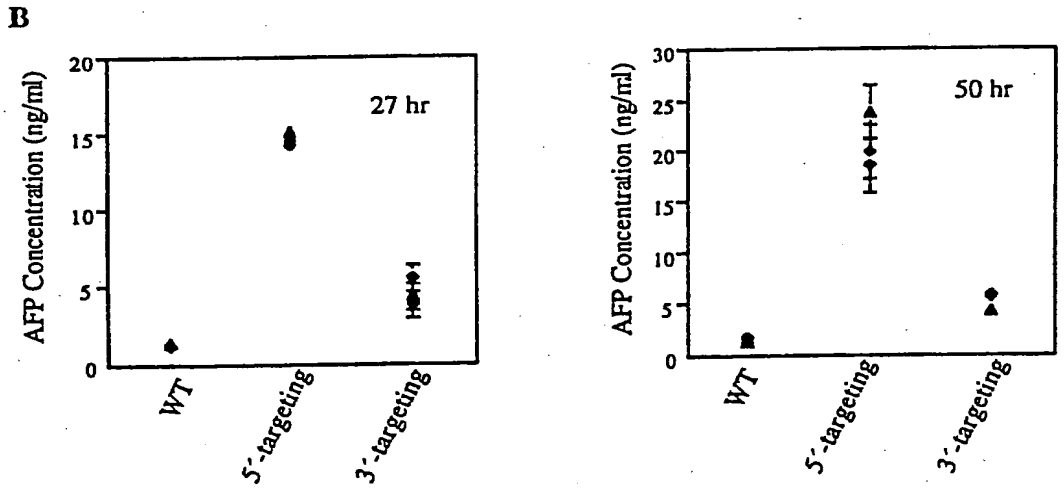

B
AFP Concentration (ng/ml)
27 hr
50 hr
WT
5′-targeting
3′-targeting

Figure 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/17493

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL : 800/3, 9, 12, 13; 435/320.1, 325; 536/23.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 800/3, 9, 12, 13; 435/320.1, 325; 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN, MEDLINE, BIOSIS, EMABASE, SCISEARCH, CAPLUS, WEST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TINSLEY et al. Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene. Nature. 28 November 1996, Vol. 384, pages 349-353, whole text.	1-40
A,P	US 6,066,725 A (DEBOER et al) 23 May 2000 (23.05.00), see entire document.	
Y	US 5,633,076 A (DEBOER et al) 27 May 1997 (27.05.97), see entire document, especially columns 1-20.	1-40
Y	US 5,850,003 A (MCLONLOGUE et al) 15 December 1998 (15.12.98), see entire document.	1-40

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 AUGUST 2000	Date of mailing of the international search report 31 AUG 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer PETER PARAS JR Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/17493

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,272,072 A (KANEKO et al) 21 December 1993 (21.12.93), see entire document.	1-40
Y	US 5,538,877 A (LUNDQUIST et al) 23 July 1996 (23.07.96), see entire document.	1-40

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/17493

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

G01N 33/00; A01K 67/00, 67/033; C12N 15/00, 15/09, 15/63, 15/70, 15/74, 5/00, 5/02; C07H 21/02, 21/04